STATISTICAL AND ENGINEERING METHODS FOR MODEL ENHANCEMENT

A Thesis
Presented to
The Academic Faculty

by

Chia-Jung Chang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2012

STATISTICAL AND ENGINEERING METHODS FOR MODEL ENHANCEMENT

Approved by:

Professor Roshan Joseph Vengazhiyil, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Nicoleta Serban
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Jianjun Shi, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor C. F. Jeff Wu
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Shreyes N. Melkote
George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: 27 April 2012

To my parents,

for their love, endless support, and encouragement.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my two advisors, Professor Jianjun Shi and Professor Roshan Joseph Vengazhiyil, for their devoted guidance, insightful advice, and unflinching encouragement throughout my doctoral study. I am particularly grateful for the incredible patience, the flexibility, and the trust they have provided when I explored the new research directions. Without their tremendous support, this dissertation would not have been existed. Being their student, working with and learning from them is truly my pleasure and such a memorable and enjoyable experience for me. Their kindness, enthusiasm, and wisdom in both research and personal life have great influence on me which will be everlasting.

I would like to thank Professor C. F. Jeff Wu, who provided me the opportunity to study at Georgia Institute of Technology. His encouragement and numerous suggestions enlightened my academic life significantly. I am especially lucky to have Professor Shreyes N. Melkote and Professor Nicoleta Serban as my other committee members, for their constructive discussions to improve my dissertation, and for their considerations and assistance during my doctoral study.

Special thanks go to Professor Sheng-Tsaing Tseng and Professor Wheyming Tina Song, who, as my advisors during my master study at Tsing-Hua University in Taiwan, introduced quality control and industrial statistics to me. It was Professor Sheng-Tsaing Tseng who recommended me to pursue Ph.D. study at Georgia Institute of Technology and encouraged me during my doctoral study.

In addition to my advisors and committee members, there are many researchers and collaborators who have shaped my academic life in various ways. In particular, I would like to thank Professor Qiang Huang and Professor Chuck Zhang, for their

collaboration on several rewarding projects and the opportunities they have provided. I would also like to thank Professor Nagi Gebraeel, for his kind support and assistance throughout my doctoral study and job search. I would also like to thank Dr. Andrew Hunt, Dr. Ganesh Venugopal, and all their colleagues in *n*Gimat Co., for their technical support on data collection.

There are many people whose friendship, moral support, and warm encouragement made my study at Atlanta memorable and colorful. I am indebted to Seonghye Jeon and Hyojung Kang, for always keeping their doors open to me, for their thoughtful considerations, and especially for their warm support in difficult times, and wonderful company in good times. A special gratitude goes to Mrs. Shi for her warm encouragement and kind consideration to my personal development; I am very fortunate to know her in Atlanta. I gratefully thank all my colleagues, friends, and visiting scholars at Georgia Institute of Technology for their cheerful presence, which made Atlanta more like home. Thank Shan Ba, Xinwei Deng, David Gass, Li Hao, Yan Hao, Ya-Lin Huang, Yan-Yu Huang, Ying Hung, Ran Jin, Lulu Kang, Kaibo Liu, Chien-Yu Peng, Matthew Plumlee, Fei Sun, Yijie Wang, Huizhi Xie, Liang Ye, Weidong Zhang, Hongxu Zhao and Cheng Zhou and many others whom have shared my life.

Last but not least, I would like to thank my parents, my brother, and my sister for their unflinching faith and endless support during my life all along. I am fortunate to have their love which gives me the courage and strength to overcome the difficulties.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Models which describe the performance of physical process are essential for quality prediction, experimental planning, process control and optimization. Engineering models developed based on the underlying physics/mechanics of the process such as analytic models or finite element models are widely used to capture the deterministic trend of the process. However, there usually exists stochastic randomness in the system which may introduce the discrepancy between physics-based model predictions and observations in reality. Alternatively, statistical models can be used to develop models to obtain predictions purely based on the data generated from the process. However, such models tend to perform poorly when predictions are made away from the observed data points.

This dissertation contributes to model enhancement research by integrating physics-based model and statistical model to mitigate the individual drawbacks and provide models with better accuracy by combining the strengths of both models. The proposed model enhancement methodologies including the following two streams: (1) data-driven enhancement approach and (2) engineering-driven enhancement approach. Through these efforts, more adequate models are obtained, which leads to better performance in system forecasting, process monitoring and decision optimization.

Among different data-driven enhancement approaches, Gaussian Process (GP) model provides a powerful methodology for calibrating a physical model in the presence of model uncertainties. However, if the data contain systematic experimental errors, the GP model can lead to an unnecessarily complex adjustment of the physical

model. In Chapter 2, we proposed a novel enhancement procedure, named as "Minimal Adjustment", which brings the physical model closer to the data by making minimal changes to it. This is achieved by approximating the GP model by a linear regression model and then applying a simultaneous variable selection of the model and experimental bias terms. Two real examples and simulations are presented to demonstrate the advantages of the proposed approach.

Different from enhancing the model based on data-driven perspective, an alternative approach is to focus on adjusting the model by incorporating the additional domain or engineering knowledge when available. This often leads to models that are very simple and easy to interpret. The concepts of engineering-driven enhancement are carried out through two applications to demonstrate the proposed methodologies.

In the first application where polymer composite quality is focused, nanoparticle dispersion has been identified as a crucial factor affecting the mechanical properties. Transmission Electron Microscopy (TEM) images are commonly used to represent nanoparticle dispersion without further quantifications on its characteristics. In Chapter 3, we developed the engineering-driven nonhomogeneous Poisson random field modeling strategy to characterize nanoparticle dispersion status of nanocomposite polymer, which quantitatively represents the nanomaterial quality presented through image data. The model parameters are estimated through the Bayesian MCMC technique to overcome the challenge of limited amount of accessible data due to the time consuming sampling schemes.

The second application is to calibrate the engineering-driven force models of laser-assisted micro milling (LAMM) process statistically, which facilitates a systematic understanding and optimization of targeted processes. In Chapter 4, the force prediction interval has been derived by incorporating the variability in the runout parameters as well as the variability in the measured cutting forces. The experimental results indicate that the model predicts the cutting force profile with good accuracy using a

95% confidence interval.

To conclude, this dissertation is the research drawing attention to model enhancement, which has considerable impacts on modeling, design, and optimization of various processes and systems. The fundamental methodologies of model enhancement are developed and further applied to various applications. These research activities developed engineering compliant models for adequate system predictions based on observational data with complex variable relationships and uncertainty, which facilitate process planning, monitoring, and real-time control.

CHAPTER I

INTRODUCTION

1.1 Problem Statement and State-of-the-art of Model Development and Model Enhancement

Models which describe the performance of physical process are essential for quality prediction, experimental planning, process control and optimization. One approach to develop models are based on the engineering knowledge and/or physics laws which govern the process. Analytical models and finite element models (FEM) are commonly used to develop such kind of physics-based models. Note that people use the two terms engineering-based model and physics-based model exchangeably in this field. Although the developments of such kind of models have been commonly addressed in various engineering areas and are quite successful, one main disadvantage of physics-based model is that the model predictions may not be accurate enough compared to true measurements collected from experimentation or production. One reason is that there usually exists stochastic randomness in the system while the physics-based models are invented to capture the mean trend of the response, which may not be able to take the stochastic portion into account. Moreover, often times there are unknown parameters in the physics-based model which need to be estimated. Researchers may know the possible ranges of these parameters but not the exact values of them. If the values of unknown parameters adopted in physics-based model are not accurate, the resultant prediction error will be large. Besides, the discrepancy may also be caused due to the non-confirming assumptions made when deriving the physics-based models, where these assumptions may not always hold in practice.

An alternative way to make predictions to better understand system performance is to borrow the statistics techniques to postulate models based on the data generated from the real process. This type of statistical models are popular when performing process quality improvement (see Wu and Hamada (2000)). One nice property of the statistical models is that it can provide good predictions when the testing points are close to the observed training data. However, it may have poor prediction accuracy

and sometimes the predictions make no physical sense especially when the points are far away from the modeling data. Moreover, the required data amount from experimentation for developing such statistical models can be quite expensive.

Many researchers have been working on how to improve the drawbacks of the aforementioned models to provide new model with better predictive accuracy. Two main streams of model enhancement works include: (1) adjusting the model by postulating statistical models on the observed prediction discrepancy, and (2) incorporating additional domain or engineering knowledge to improve the originally existing models. The state-of-the-art developments with respect to each stream will be reviewed individually as follows.

The idea of improving physics-based model based on the experimental data is not new. One way to do so is to estimate the unknown model parameters (which is also known as calibration parameters) in the physics-based model with the data collected from real experimentation or production. This type of model is the so called mechanistic model, see Box and Hunter (1962), Kapoor et al. (1998) for more examples. However, when there is no calibration parameter in the physics-based model, this mechanistic-modeling method can not be used. In addition, this type of model does not consider model inadequacy and thus is not able to serve model correction purpose.

Kennedy and O'Hagan (2001) proposed a model which capture the discrepancy between the physics-based model and real data by using a Gaussian Process (GP). The proposed model calibration idea takes model inadequacy into account. More works related to their modeling approach are proposed by Higdon et al. (2004), Bayarri et al. (2007), Qian and Wu (2008), Joseph and Melkote (2009), Wang, Chen, and Tsui (2009), Han, Santner, and Rawlinson (2009), among many others. A recent review of this important topic can be found in Xiong et al. (2009).

Instead of improving the physics-based model by capturing the discrepancy between model predictions and real observations with GP in data-driven sense, another approach to improve model predictive ability is to incorporate additional domain/engineering knowledge into model development directly. This kind of model enhancement concept has been commonly adopted in various mechanical engineering applications. It has been demonstrated that the original physics-based model could

be improved by using the addition of other developed engineering knowledge.

For example, Pan et al. (2009) proposed an engineering-driven rule-based detection (ERD) method to develop an on-line quality inspection algorithm for detecting the surface defect ("bleed") generated in continuous casting processes. The ERD consists of three detection stages using the pixel features of bleeds, which are transferred from the physical features generated via engineering knowledge. Another example is an engineering-driven reconfiguration method for piecewise linear regression trees proposed by Jin and Shi (2011). The authors take available engineering knowledge into account so that a feasible feed-forward control based on piecewise linear models are developed to model the variation and its propagation from those observational data. Moreover, in Zhao et al. (2011), a PDE-constrained Gaussian Process model is developed based on Global Galerkin discretization of the governing Partial Differential Equations (PDEs). In this example, the important ingredients for model development is the PDEs, which govern the wire saw slicing process, are obtained from engineering knowledge. Real experiments are conducted to provide data for the validation of the proposed model in this work which demonstrates the effectiveness of improving physics-based model by means of existing engineering knowledge.

1.2 Research Objectives and Organization of the Dissertation

In this dissertation, we aim to integrate physics-based model and statistical model to mitigate their individual drawbacks and provide models with better accuracy by combining the strengths of the two models. The proposed enhanced models are expected to produce more realistic predictions than both engineering models and statistical models alone while computationally tractable in terms of model development and parameter estimation efforts.

This dissertation consists of three papers accepted/submitted for journal publications. These papers are on various aspects of model enhancement methodologies in complex manufacturing processes. Figure 1 outlines the structure of this dissertation along with the focus of each chapter individually.

The overview of the concepts and state-of-the-art of model development and model

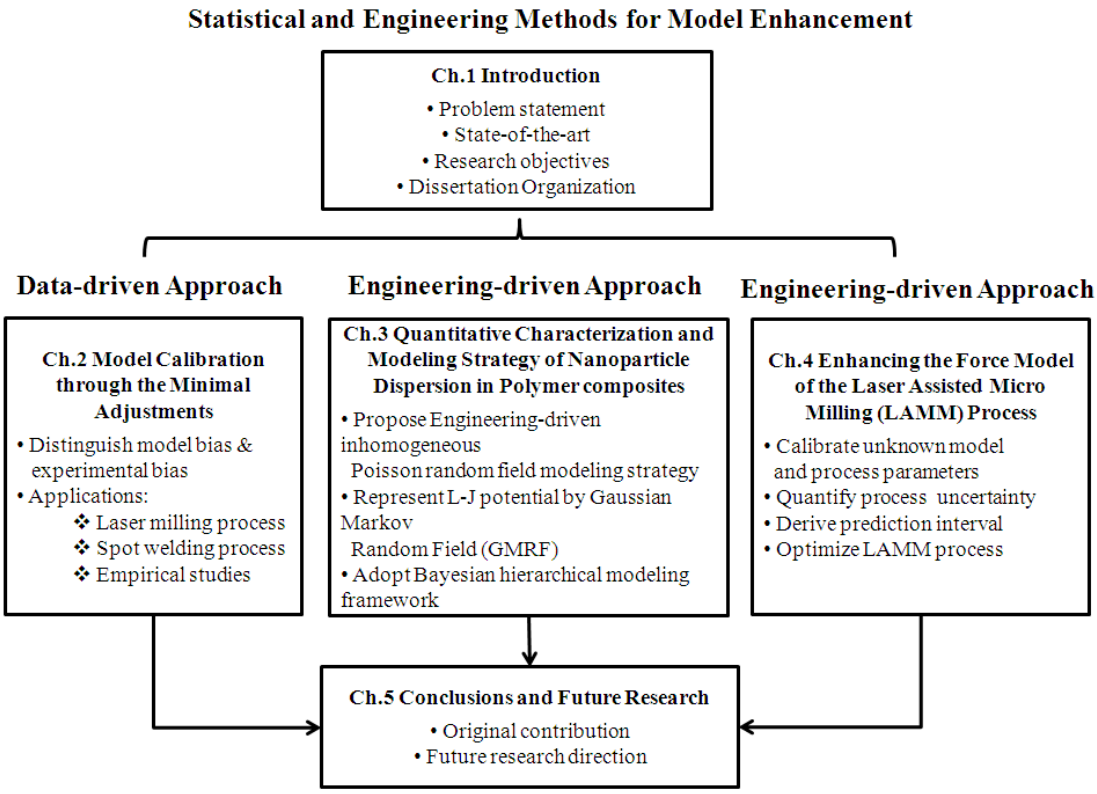


Figure 1: Outline of the dissertation

enhancement has been discussed in Chapter 1. In Chapter 2, the data-driven adjustment methodology which enhances the original physics-based model based on the observed data discrepancy by statistical methods will be discussed in details. The proposed "minimal adjustment approach" not only develops the model with better predictive accuracy but also provides the potential for the interpretation of model discrepancy from the physics or engineering knowledge sense. The methodological development of the proposed approach requires the techniques such as linear approximation of GP, variable selection, surrogate model development and so on. The proposed approach has been applied to laser drilling process, spot welding process, and empirical simulation study to demonstrate the outperformed capability of model prediction.

Afterwards, the emphasis will be moved to the topics on how to incorporate the additional engineering knowledge when performing model enhancement to obtain models with better accuracy. The concepts and detailed procedures are discussed through two applications: (1) quantification of nanoparticle dispersion, see Chapter 3, and (2)

force model development and enhancement in LAMM process, see Chapter 4.

In Chapter 3, the quantification method for nanoparticle dispersion is presented. Polymer nanocomposites nowadays have been received tremendous attention due to the outperformed mechanical properties. It has been studied that the nanoparticle dispersion plays a crucial role in the mechanical properties of polymer nanocomposites. However, how to quantify the nanoparticle dispersion state based on the Transmission Electron Microscopy (TEM) images taken from polymer nanocomposite samples are an challenging task with a strong need. The objective of Chapter 3 is to develop an effective modeling strategy to characterize nanopartial dispersion among different locations of a nanocomposite surface. The proposed method combines the production model consisted by process variables and engineering knowledge capturing the nanoparticle interaction phenomenon together and represent the nanoparticle dispersion states at nano-scale by means of an engineering-driven nonhomogeneous Poisson random field. The model parameters are estimated through the Bayesian MCMC technique to overcome the challenge of limited amount of accessible data due to the time consuming sample collection process. The TEM images taken from nano-silica/epoxy composites will be used to support the proposed methodology. The research strategy and framework are generally applicable to other nanocomposite materials.

In Chapter 4, the research objective is to develop the force prediction model in Laser Assisted Micro Milling (LAMM) process. LAMM is capable of generating three dimensional micro-scale features in hard metals with reduced cutting forces compared to conventional micro milling. To maximize this reduction in the cutting forces, a mathematical model is required to understand the influence of different laser and machining parameters on the cutting forces, in addition to the estimation of temperature distribution in LAMM. The computer experimentation to better understand temperature change in LAMM process is available by finite-element modeling (FEM) while the computational time is quite expensive. As a result, the surrogate model is developed to capture the temperature change with different laser settings. Finally, a physics-based force model with engineering adjustment is developed to predict cutting forces when machining a hard metal using laser assist. The force prediction interval has been derived by incorporating the variability in the runout parameters as well

as the variability in the measured cutting forces. Moreover, LAMM experiments are carried out on 52100 bearing steel (62 HRc), over a range of feed rates and laser powers and the resultant cutting forces have been collected. The experimentation results indicate that the proposed force model predicts the cutting force profile with good accuracy using using a 95% confidence interval.

Finally, Chapter 5 concludes the dissertation, summarizes the original contributions and discusses the future research directions.

1.3 References

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C. H., Tu, J. (2007), "A Framework for Validation of Computer Models," *Technometrics*, 49, 138–154.

Box, G. E. P. and Hunter, W. G. (1962). "A Useful Method for Model-Building". *Technometrics* 4, pp. 301318.

Han, G., Santner, T. J., and Rawlinson, J. J. (2009), "Simultaneous Determination of Tuning and Calibration Parameters for Computer Experiments," *Technometrics*, 51, 464-474.

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004), "Combining Field Data and Computer Simulations for Calibration and Prediction," *SIAM Journal of Scientific Computing*, 26, 448–466.

Jin, R. and Shi, J., 2011, "Reconfigured Piecewise Linear Regression Tree for Multistage Manufacturing Process Control," *IIE Transactions*, in-press.

Joseph, V. R. and Melkote, S. N. (2009), "Statistical Adjustments to Engineering Models," *Journal of Quality Technology*, 41, 362–375.

Pan, E., Ye, L., Shi, J., Chang, T. -S, (2009), "On-line bleeds detection in continuous casting processes using engineering-driven rule-based algorithm", *ASME Transactions, Journal of Manufacturing Science and Engineering*, Volume 131, Issue 6.

Qian, Z., Wu, H. and Wu, C. F. J. (2008), "Gaussian Process Models for Computer Experiments with Qualitative and Quantitative Factors," *Technometrics*, 50, 383–396.

Wang, S., Chen, W., and Tsui, K.-L. (2009), "Bayesian Validation of Computer Models," *Technometrics*, 51, 439–451.

Wu, C. F. J. and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York, NY: Wiley.

Xiong, Y., Chen, W., Tsui, K.-L., and Apley, D. (2009), "A better understanding of model updating strategies in validating engineering models," *Computational Methods in Applied Mechanical Engineering*, 198, 1327–1337.

Zhao, H., Jin, R., Wu, S. and Shi, J., 2011, "PDE-constrained Gaussian Process Model on Material Removal Rate of Wire Saw Slicing Process," *ASME Transactions, Journal of Manufacturing Sciences and Engineering*, Vol. 133, 2.

CHAPTER II

MODEL CALIBRATION THROUGH THE MINIMAL ADJUSTMENTS

2.1 Introduction

A physics-based model often contains unknown parameters. Model calibration refers to estimating these unknown parameters from real data. Traditionally, least squares methods are often employed for this purpose. A major assumption underlying in this approach is that the physical model is an accurate representation of the system. When this assumption is violated, the estimates of the unknown *calibration parameters* become inaccurate leading to poor model prediction.

Kennedy and O'Hagan (2001), in a seminal paper, addressed this problem by adding a bias term to the model to capture the discrepancy between the physical model and the "true" model. Their approach can be described as follows. Let y be the physical characteristic of the system which we are interested in predicting with respect to the variables $\boldsymbol{x} = (x_1, \cdots, x_p)'$. Let $f(\boldsymbol{x}; \boldsymbol{\eta})$ be the physical model, where $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_q)'$ are the unknown calibration parameters. Then, an observation from the system is modeled as

$$y = \mu(\boldsymbol{x}) + \epsilon, \tag{1}$$

where the true model $\mu(\boldsymbol{x})$ is corrupted by a random noise $\epsilon \sim N(0, \sigma^2)$. Now, add the bias $m(\boldsymbol{x})$ to the physical model, i.e.,

$$\mu(\boldsymbol{x}) = \rho f(\boldsymbol{x}; \boldsymbol{\eta}) + m(\boldsymbol{x}), \tag{2}$$

where ρ is an additional term inserted into the model for capturing the scale bias. It is expected that if the model bias terms are accurately estimated, then it will lead to better estimates of the calibration parameters and subsequently improved predictions. Kennedy and O'Hagan's approach gives a broader meaning for model calibration as bringing the model closer to reality. Further advancement of their modeling approach was made by Higdon et al. (2004), Bayarri et al. (2007), Qian and Wu (2008), Joseph

and Melkote (2009), Wang, Chen, and Tsui (2009), Han, Santner, and Rawlinson (2009), among many others. A recent review of this important topic can be found in Xiong et al. (2009).

The foregoing work have not entertained another potentially important bias that can occur, not in the model, but in the observations. The observations are usually generated from the system by conducting experiments. The nature of the experiment can create certain errors such as set-up errors, which although occur randomly during experiment can cause systematic biases in model estimation. For example, Deng et al. (2009) studied the estimation of elastic modulus of a nanobelt using a simply supported beam model. The observations, however, were corrupted by sample manipulation and stick-slip events which occurred randomly during the measurement process. Note that these errors are only experimental and not related to the model discrepancy.

Experimental bias can cause serious errors in model estimation. A simple illustrative example is shown in Figure 2. The true relationship between x and y is plotted as a solid line. For simplicity take the physical model to be the same as the true model. The experiment is conducted at five different values of x and 10 replicates are collected at each x. Suppose a set up error occurred at $x = 30$, which shifted the observations by about 3 units upwards. Now if we ignore this error and use Kennedy and O'Hagan's approach, then the estimated model will be as shown as the dashed line. It can be seen that the estimated model is quite different from the true model and in fact, it is made much more complex than the true model due to the presence of the experimental bias. However, we could easily identify this because we knew the true model. In reality, the experimental bias and the model bias are completely confounded and thus, it is not easy to separate them. Too see this, write the model as follows

$$y_{ij} = \rho f(x_i) + m(x_i) + e_i + \epsilon_{ij},$$

for $i = 1, \cdots, 5$ and $j = 1, \cdots, 10$, where e_i denotes the experimental bias at x_i. It can be seen that $m(x_i)$ and e_i are confounded and cannot be directly identified from the data.

In this work, we develop a methodology for calibrating the physical model in the presence of both model and experimental biases. We should emphasize that the

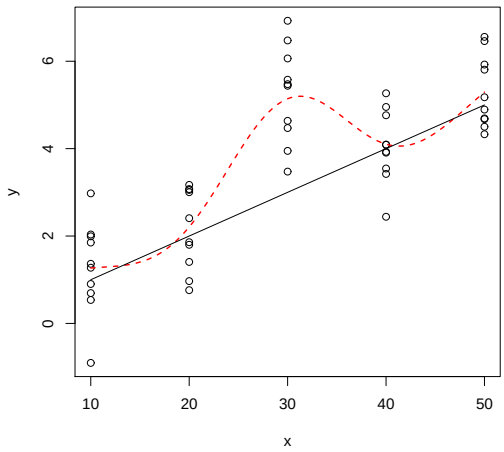


Figure 2: Effect of experimental bias on model estimation: true model (solid) and estimated model (dashed).

introduction of experimental bias is an important twist to the problem formulation because experimental bias is almost always present in physical experiments and occur more often than the model bias. Separating experimental bias from model bias seems like a hopeless endeavor given the complete confounding between the two biases. Nevertheless, we show that the nature of these biases can be used to separate them successfully, at least in some cases. The main idea of the methodology is explained in Section 2 and the estimation details are given in Section 3. The methodology is illustrated with two real examples in Section 4. Some simulations are performed in Section 5 to assess the effectiveness of the proposed methodology and we conclude with some remarks in Section 6.

2.2 Methodology

Suppose that the physical experiment is performed at n different settings (or locations) $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ and r replicates are collected at each of the settings. Let y_{ij} be the jth replicate observed at the ith setting. To reduce the cost and time, the experiment is usually performed by setting the $\boldsymbol{x}$ variables at some value and then collecting all of the replicates simultaneously. This introduces a common experimental bias term, e_i, for all the replicates within an experimental setting. Then, the model becomes

$$y_{ij} = \rho f_i(\boldsymbol{\eta}) + m_i + e_i + \epsilon_{ij}, \tag{3}$$

where $f(\boldsymbol{x}_i;\boldsymbol{\eta}) = f_i(\boldsymbol{\eta})$ and $m(\boldsymbol{x}_i) = m_i$ for $i = 1, \cdots, n$. We can see that m_i and e_i are completely confounded.

It is important to note that even when the experimental design is completely randomized, the experimental bias term can be present due to the finite number of replicates. An extreme but common case is an unreplicated experiment, where we do not have any replicates to randomize. A complete randomization can reduce the experimental bias but not eliminate it. Thus, the foregoing model will be useful irrespective of the type of randomization employed in the experiment. This will be shown in the next section.

The objective is to estimate $\boldsymbol{\eta}$, ρ, and $m(\boldsymbol{x})$. Clearly, the confounding problem is not limited to the model bias and experimental bias. For any value of ρ or $\boldsymbol{\eta}$, we can choose m_i, e_i, and ϵ_{ij} to get the same y_{ij}. Thus there is a confounding among ρ, $\boldsymbol{\eta}$, e_i, m_i, and ϵ_{ij}. In the original formulation of the problem, Kennedy and O'Hagan (2001) also assumed that the physical model is a complex computer model and hence not known for all $\boldsymbol{x}$. This introduces another level of confounding.

To mitigate the confounding among some of the terms, Bayarri et al. (2007) suggested to fix $\rho = 1$ and use modularization in estimation, i.e., only the computer model outputs are used for estimating $f(\boldsymbol{x})$. Further justification to this approach was given by Liu, Bayarri, and Berger (2009). They also suggested to estimate σ^2 using only the replicates (note that $\epsilon_{ij} \sim N(0, \sigma^2)$). However, because the data cannot be divided into two separate modules for estimating the experimental and model biases, modularization cannot be applied here. Thus, a different framework is needed for separating the experimental and model biases.

Our approach can be explained as follows. First, as done in Joseph and Melkote (2009), reparameterize the model as

$$y_{ij} - f_i(\boldsymbol{\eta}) = \alpha_0 + \alpha_1(f_i(\boldsymbol{\eta}) - \bar{f}(\boldsymbol{\eta})) + m_i + e_i + \epsilon_{ij}, \tag{4}$$

where $\bar{f}(\boldsymbol{\eta}) = 1/n \sum_{i=1}^{n} f_i(\boldsymbol{\eta})$. A crucial idea to separate experimental and model bias is to recognize that the model bias is a continuous and smooth function in $\boldsymbol{x}$, whereas the experimental biases are discrete and present only at the experimental locations. Therefore, we expand the model bias using some continuous basis functions:

$$m(\boldsymbol{x}) = \sum_{k=0}^{N} \beta_k u_k(\boldsymbol{x}),$$

where $\{u_0(\boldsymbol{x}), \cdots, u_N(\boldsymbol{x})\}$ is a set of known functions and β_i's are unknown parameters. In this work, we use the Legendre orthogonal polynomial basis functions. Other bases such as wavelets and Fourier series can also be used depending on the type of problem. First, we scale all the variables in $[-1, 1]$. Then, a qth degree Legendre polynomial (normalized to unit length) of a particular variable x is given by (see, e.g., Hildebrand 1987, page 329)

$$P_q(x) = \frac{\sqrt{q+1/2}}{2^q q} \frac{d^q}{dx^q}(x^2-1)^q.$$

It is easy to see that $P_0(\boldsymbol{x}) = 1$, $P_1(x) = x$, $P_2(x) = \sqrt{5/8}(3x^2-1)$, $P_3(x) = \sqrt{7/2}(5x^3-3x)$, $\cdots$, which can be used to define the overall mean, linear, quadratic, cubic effects, etc. of the variable x. The Legendre polynomials of two or more variables can be multiplied together to define their interaction effect. The degree of Legendre polynomials for each variable and the order of interaction effects among the variables can be chosen depending on the experimental design. They should be chosen so that the total number of basis functions $(N+1)$ is larger than n. This ensures that the size of the basis increases with the size of the data. In other words, more complex model bias terms are entertained when there is more data.

Thus, the model becomes

$$y_{ij} - f_i(\boldsymbol{\eta}) = \alpha_0 + \alpha_1(f_i(\boldsymbol{\eta}) - \bar{f}(\boldsymbol{\eta})) + \sum_{k=0}^{N} \beta_k u_k(\boldsymbol{x}_i) + e_i + \epsilon_{ij}. \tag{5}$$

At first sight, we might think that both α_0 and β_0 are not required, but as will be seen later, using two separate constants facilitate the estimation procedure. This is an overparameterized model with unknown parameters $\{\boldsymbol{\eta}, \alpha_0, \alpha_1, \beta_0 \cdots, \beta_N, e_1, \cdots, e_n\}$. Therefore, we use a variable selection technique to identify and estimate the important (or significant) parameters.

Different strategies can be adopted for variable selection such as least angle regression (Efron et al. 2004) and lasso (Tibshirani 1996). Here we use a strategy based on the nonnegative (nn) garrote (Brieman 1995) because of its flexibility for accommodating hierarchical relationships among predictors and ease of implementation. Moreover, as explained in the next section, it helps in making a connection with the existing model calibration approach of Kennedy and O'Hagan (2001). The nn-garrote works by first obtaining a set of base estimates (sometimes called initial

estimates) and then the final estimates using a constrained least squares optimization. Let $\tilde{\alpha}_0,\ \tilde{\alpha}_1,\ \tilde{\beta}_0, \cdots, \tilde{\beta}_N,\ \tilde{e}_1, \cdots, \tilde{e}_n$ be the base estimates of the parameters. Let $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_{N+n+2})'$ be the nonnegative parameters used in the nn-garrote method. Then, $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are estimated by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{r}\left\{y_{ij} - f_i(\boldsymbol{\eta}) - \theta_0\tilde{\alpha}_0 - \theta_1\tilde{\alpha}_1(f_i(\boldsymbol{\eta}) - \bar{f}(\boldsymbol{\eta})) - \sum_{k=0}^{N}\theta_{k+2}\tilde{\beta}_k u_k(\boldsymbol{x}_i) - \theta_{N+2+i}\tilde{e}_i\right\}^2, \tag{6}$$

subject to the constraints

$$\begin{aligned}\sum_{k=0}^{N+n+2}\theta_k &\leq M,\\ \theta_k &\geq 0 \text{ for all } k = 0, \cdots, N+n+2.\end{aligned}$$

When $M = 0$, all of the θ's will be 0 and thus, none of the variables will be selected. As M increases, some of the θ's become positive thereby introducing the corresponding variables into the model. Note that in this formulation of the problem there are no θ parameters to control the values of the calibration parameters. Thus, $\boldsymbol{\eta}$ will be chosen to minimize the sum of squares of errors and then, the model and experimental bias terms are added, only if they are needed.

The simultaneous variable selection employed in the nn-garrote method is another crucial idea to separate experimental biases from model biases. If the experimental biases were not included in the variable selection procedure, then it would have wrongly classified as due to model bias and would have resulted in a complex model adjustment. Thus, by using this procedure, we can identify the minimum number of terms to be included in the model bias and therefore, we call this as minimal adjustment procedure. This is a concept initially discussed by Joseph and Melkote (2009), but was not well understood in their work.

The optimization in (6) is difficult because of the presence of $\boldsymbol{\eta}$. We simplify it as follows. Start with an initial estimate of $\boldsymbol{\eta}$, which can be taken as the estimate obtained at the previous value of M in the nn-garrote. Then, we solve the quadratic program to obtain the $\boldsymbol{\theta}$. This is easy to do. Now, we fix the $\boldsymbol{\theta}$ and estimate the $\boldsymbol{\eta}$, which can be done using nonlinear regression. These iterations are continued until convergence (the convergence is guaranteed if (6) is jointly convex with respect to $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$). In fact, not more than one iteration is needed at each value of M because

the optimization problem is solved at small increments of M and therefore, not much changes will happen to the parameters obtained at the previous value of M.

Least squares estimates are often used as base estimates for the parameters in the nn-garrote. However, least squares estimates of $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)'$, $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_N)'$, and $\boldsymbol{e} = (e_1, \cdots, e_n)'$ do not exist because of having less data than the number of parameters and therefore, alternative ways of obtaining initial values need to be investigated. This is done in the next section.

2.3 Estimation

When least squares estimates of the parameters do not exist, ridge regression can be used for obtaining the base estimates of the parameters in the nn-garrote (Yuan and Lin 2007). However, the ridge regression treats all the parameters the same and ignores the hierarchical relationships among the variables. One way to overcome this deficiency is to add additional constraints to the nn-garrote parameters to satisfy the hierarchical relationships as done in Yuan, Joseph, and Zou (2009). Another way is to choose an appropriate scale matrix in the ridge regression instead of an identity matrix. This can be done by using the prior variance-covariance matrix for the linear model parameters proposed in Joseph (2006) and Joseph and Delaney (2007). This prior variance-covariance matrix is obtained from a Gaussian Process (GP) prior. Interestingly, Kennedy and O'Hagan (2001)'s approach to model calibration uses a GP prior for the $m(\boldsymbol{x})$. Therefore, we chose the second idea of using a scale matrix in the ridge regression because it makes a nice connection between Kennedy and O'Hagan's approach and our proposed approach.

2.3.1 Base Estimates Using Gaussian Process

As in Kennedy and O'Hagan (2001) assume a GP model for the model bias:

$$m(\boldsymbol{x}) \sim GP(0, \tau^2\psi),$$

where the correlation function $\psi(\boldsymbol{x}, \boldsymbol{t})$ is defined as $cor\{m(\boldsymbol{x}), m(\boldsymbol{t})\}$. A common choice for the correlation function is the Gaussian correlation function given by

$$\psi(\boldsymbol{x}, \boldsymbol{t}) = \exp\{-\sum_{i=1}^{p} c_i(x_i - t_i)^2\},$$

where $\boldsymbol{c} = (c_1, \cdots, c_p)'$ are unknown correlation parameters. In a Bayesian framework, the GP can be viewed as a prior distribution to the unknown function $m(\boldsymbol{x})$. Similarly, assume the following prior for the experimental bias:

$$e_i \sim^{iid} N(0, \lambda^2),$$

for $i = 1, \cdots, n$. Note that the experimental bias is not continuous like the model bias and therefore, the correlations of e_i's are taken as 0 assuming the n experimental settings are done independent of each other. If restrictions on randomization are made within the factor levels such as in split-plot experiments, then an appropriate model for e_i's that captures the correlations induced by the restricted randomization should be used.

We first estimate the $m(\boldsymbol{x})$ from data and then obtain estimates of the β's by approximating it by a linear model. For the moment assume that $\boldsymbol{\eta}$, $\boldsymbol{\alpha}$, τ^2, $\boldsymbol{c}$, λ^2, and σ^2 are known. We will explain later how they can be estimated. Let $\boldsymbol{y} = (y_{11}, \cdots, y_{n1}, \cdots, y_{1r}, \cdots, y_{nr})'$ be the data vector. Let $\boldsymbol{F}(\boldsymbol{\eta})$ be an $n \times 2$ matrix, whose first column is a column of 1's and the second column is $(f_1(\boldsymbol{\eta}) - \bar{f}(\boldsymbol{\eta}), \ldots, f_n(\boldsymbol{\eta}) - \bar{f}(\boldsymbol{\eta}))'$ and $\boldsymbol{f}(\boldsymbol{\eta}) = (f_1(\boldsymbol{\eta}), \ldots, f_n(\boldsymbol{\eta}))'$. Then, it is easy to show that the posterior mean of $m(\boldsymbol{x})$ is given by

$$\tilde{m}(\boldsymbol{x}) = \tau^2 \psi(\boldsymbol{x})' \otimes \mathbf{1}_r' \{(\tau^2 \boldsymbol{\Psi} + \lambda^2 \boldsymbol{I}_n) \otimes \boldsymbol{J}_r + \sigma^2 \boldsymbol{I}_{nr}\}^{-1} (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\eta}) \otimes \mathbf{1}_r - \boldsymbol{F}(\boldsymbol{\eta})\boldsymbol{\alpha} \otimes \mathbf{1}_r), \quad (7)$$

where $\mathbf{1}_r$ is a column of 1's having length r, $\boldsymbol{I}_r$ is an $r \times r$ identity matrix, $\boldsymbol{J}_r$ is an $r \times r$ matrix of 1's, and $\otimes$ denotes Kronecker product. Similarly, if we were to use a completely randomized design, then

$$\tilde{m}(\boldsymbol{x}) = \tau^2 \psi(\boldsymbol{x})' \otimes \mathbf{1}_r' \{\tau^2 \boldsymbol{\Psi} \otimes \boldsymbol{J}_r + (\lambda^2 + \sigma^2) \boldsymbol{I}_{nr}\}^{-1} (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\eta}) \otimes \mathbf{1}_r - \boldsymbol{F}(\boldsymbol{\eta})\boldsymbol{\alpha} \otimes \mathbf{1}_r), \quad (8)$$

The posterior mean of $\boldsymbol{e}$ can be obtained similarly. The computation of $\tilde{m}(\boldsymbol{x})$ and $\tilde{\boldsymbol{e}}$ can become difficult if nr is large. We have the following useful result to reduce the computations and simplify the results.

PROPOSITION 1: The posterior mean of $m(\boldsymbol{x})$ is given by

$$\tilde{m}(\boldsymbol{x}) = \tau^2 \psi(\boldsymbol{x})' \{\tau^2 \boldsymbol{\Psi} + \nu^2 \boldsymbol{I}_n + \frac{\sigma^2}{r} \boldsymbol{I}_n\}^{-1} (\bar{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{\eta}) - \boldsymbol{F}(\boldsymbol{\eta})\boldsymbol{\alpha}), \quad (9)$$

and the posterior mean of $\boldsymbol{e}$ is given by

$$\tilde{\boldsymbol{e}} = \nu^2 \{\tau^2 \boldsymbol{\Psi} + \nu^2 \boldsymbol{I}_n + \frac{\sigma^2}{r} \boldsymbol{I}_n\}^{-1} (\bar{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{\eta}) - \boldsymbol{F}(\boldsymbol{\eta})\boldsymbol{\alpha}), \quad (10)$$

where $\bar{\boldsymbol{y}} = (\bar{y}_{1.}, \cdots, \bar{y}_{n.})'$, $\bar{y}_{i.} = \sum_{j=1}^{r} y_{ij}/r$, $\nu^2 = \lambda^2$ when the replicates are not randomized among the experimental runs, and $\nu^2 = \lambda^2/r$ in a completely randomized design.

Proof. Let $\boldsymbol{B} = (\tau^2\boldsymbol{\Psi} + \lambda^2\boldsymbol{I}_n)/\sigma^2$. Then, by using the matrix identity $(\boldsymbol{I}_{nr} + \boldsymbol{B} \otimes \boldsymbol{J}_r)^{-1} = \boldsymbol{I}_{nr} - \boldsymbol{B}(\boldsymbol{I}_n + r\boldsymbol{B})^{-1} \otimes \boldsymbol{J}_r$, (7) can be simplified to

$$\begin{aligned}\tilde{m}(\boldsymbol{x}) &= \tau^2\psi(\boldsymbol{x})' \otimes \mathbf{1}_r' \frac{1}{\sigma^2}\{\boldsymbol{I}_{nr} - \boldsymbol{B}(\boldsymbol{I}_n + r\boldsymbol{B})^{-1} \otimes \boldsymbol{J}_r\}(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\eta}) \otimes \mathbf{1}_r - \boldsymbol{F}(\boldsymbol{\eta})\boldsymbol{\alpha} \otimes \mathbf{1}_r) \\ &= \frac{r\tau^2}{\sigma^2}\{\boldsymbol{I}_n - r\boldsymbol{B}(\boldsymbol{I}_n + r\boldsymbol{B})^{-1}\}(\bar{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{\eta}) - \boldsymbol{F}(\boldsymbol{\eta})\boldsymbol{\alpha}).\end{aligned}$$

Now by applying Woodbury's matrix inverse formula (see, e.g., Harville 1997), we obtain (9). The other results can be obtained similarly. □

Proposition 1 tells us that, we can take the average of the y_{ij}'s at each of the experimental setting and treat it as a single observation. The only difference is that we need to replace σ^2 with σ^2/r, which is expected because we are analyzing on the averages. Interestingly, the result holds for both type of randomizations. Because λ^2 is unknown in real applications and is estimated from the data, it does not matter whether we are estimating λ^2 or λ^2/r. Thus, we can replace the observations at each of the experimental setting by their average and forget about the type of randomization employed in the experiment. Of course, the randomization has an effect, which can be seen in the estimate of experimental bias. The experimental bias will be smaller in a completely randomized design compared to the restricted randomized design by an order of r.

The unknown parameters σ^2, $\boldsymbol{\eta}$, $\boldsymbol{\alpha}$, τ^2, ν^2, and $\boldsymbol{c}$ can be estimated as follows. First, as suggested by Liu et al. (2009), we estimate σ^2 from the replicates. The prior information about σ^2 can also be incorporated into this estimation (see Joseph and Melkote 2009). A reasonable estimate of the calibration parameters can be obtained by using least squares estimation assuming no model and experimental biases, which coincides with the estimate of $\boldsymbol{\eta}$ in the nn-garrote procedure at $M = 0$. Thus, let

$$\tilde{\boldsymbol{\eta}} = arg\min_{\boldsymbol{\eta}} (\bar{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{\eta}))'(\bar{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{\eta})) \tag{11}$$

Then, we obtain the estimates of the remaining parameters by maximizing the integrated likelihood, which is equivalent to minimizing

$$\log det(\boldsymbol{A}) + (\bar{\boldsymbol{y}} - \boldsymbol{f}(\tilde{\boldsymbol{\eta}}) - \boldsymbol{F}(\tilde{\boldsymbol{\eta}})\tilde{\boldsymbol{\alpha}})'\boldsymbol{A}^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{f}(\tilde{\boldsymbol{\eta}}) - \boldsymbol{F}(\tilde{\boldsymbol{\eta}})\tilde{\boldsymbol{\alpha}}), \tag{12}$$

with respect to τ^2, ν^2, and $\boldsymbol{c}$, where $\boldsymbol{A} = \tau^2\boldsymbol{\Psi} + \nu^2\boldsymbol{I}_n + \frac{\sigma^2}{r}\boldsymbol{I}_n$ and

$$\tilde{\boldsymbol{\alpha}} = \{\boldsymbol{F}(\tilde{\boldsymbol{\eta}})'\boldsymbol{A}^{-1}\boldsymbol{F}(\tilde{\boldsymbol{\eta}})\}^{-1}\boldsymbol{F}(\tilde{\boldsymbol{\eta}})'\boldsymbol{A}^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{f}(\tilde{\boldsymbol{\eta}})). \tag{13}$$

Let $\tilde{\tau}^2$, $\tilde{\nu}^2$, and $\tilde{\boldsymbol{c}}$ denote the solution. Substituting these estimates in (9) and (10), we can get base estimates of $m(\boldsymbol{x})$ and $\boldsymbol{e}$. The base estimates of $\boldsymbol{\beta}$ can now be obtained by approximating $m(\boldsymbol{x})$ by a linear model, which is explained in the next subsection.

2.3.2 Approximation of a GP model Using a Linear Regression Model

Instead of approximating the posterior distribution of $m(\boldsymbol{x})$, we will work on approximating the prior distribution. As shown in Joseph (2006), such an approach will produce much simpler results than directly approximating the posterior. Thus, our objective is to find a prior for $\boldsymbol{\beta}$ so that the distribution of $\sum_{i=0}^{N}\beta_i u_i(\boldsymbol{x})$ can approximate the GP prior.

Let $\boldsymbol{u}(\boldsymbol{x}) = (u_0(\boldsymbol{x}), \cdots, u_N(\boldsymbol{x}))'$. Now, for every realization of $m(\boldsymbol{x})$ in $\mathcal{X}$, we find a realization of $\boldsymbol{\beta}$ such that $\boldsymbol{u}(\boldsymbol{x})'\boldsymbol{\beta} \approx m(x)$. We use the least squares criterion. Thus, $\boldsymbol{\beta}$ is obtained by minimizing

$$\int_{\mathcal{X}}\{m(\boldsymbol{x}) - \boldsymbol{u}(\boldsymbol{x})'\boldsymbol{\beta}\}^2 d\boldsymbol{x},$$

where $\mathcal{X}$ denotes the experimental region $[-1, 1]^p$. The solution to this can be easily obtained as

$$\boldsymbol{\beta} = \left\{\int_{\mathcal{X}}\boldsymbol{u}(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})'d\boldsymbol{x}\right\}^{-1}\int_{\mathcal{X}} m(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})d\boldsymbol{x}. \tag{14}$$

Thus, we obtain the following important result which has applications beyond the model calibration problem discussed in this work.

THEOREM 1: Consider a Gaussian Process model $GP(0, \tau^2\psi)$ with continuous correlation function $\psi(\boldsymbol{x}, \boldsymbol{t})$ and a linear model $\boldsymbol{u}(\boldsymbol{x})'\boldsymbol{\beta}$ with each basis function $u_i(\boldsymbol{x})$ Riemann integrable. Then, the least squares approximation of the Gaussian Process model by the linear model can be obtained by setting $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \tau^2\boldsymbol{\Sigma})$, where $\boldsymbol{0}$ is a vector of 0's having length $N + 1$ and

$$\boldsymbol{\Sigma} = \left\{\int_{\mathcal{X}}\boldsymbol{u}(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})'d\boldsymbol{x}\right\}^{-1}\int_{\mathcal{X}}\int_{\mathcal{X}}\psi(\boldsymbol{x}, \boldsymbol{t})\boldsymbol{u}(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{t})'d\boldsymbol{x}d\boldsymbol{t}\left\{\int_{\mathcal{X}}\boldsymbol{u}(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x})'d\boldsymbol{x}\right\}^{-1}.$$

The following result immediately follows from Theorem 1 and the following property of Legendre polynomials:

$$\int_{\mathcal{X}} u_i(\boldsymbol{x})u_j(\boldsymbol{x})d\boldsymbol{x} \;\; = \;\; \delta_{ij},$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Proof. Because $\psi(\boldsymbol{x}, \boldsymbol{t})$ is continuous and $u_i(\boldsymbol{x})$ is Reimann integrable, the stochastic integral on the right side of (14) is well defined (Cramér and Leadbetter 1967, Section 5.3). Now, taking expectations on both sides of (14) and applying Fubini's theorem,

$$E(\boldsymbol{\beta}) = \left\{ \int_{\mathcal{X}} \boldsymbol{u}(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x})' d\boldsymbol{x} \right\}^{-1} \int_{\mathcal{X}} E(m(\boldsymbol{x})) \boldsymbol{u}(\boldsymbol{x}) d\boldsymbol{x} = \boldsymbol{0}.$$

Thus,

$$\begin{aligned}
&var(\boldsymbol{\beta}) = E(\boldsymbol{\beta}\boldsymbol{\beta}') \\
&= E\left\{ \left\{ \int_{\mathcal{X}} \boldsymbol{u}(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x})' d\boldsymbol{x} \right\}^{-1} \int_{\mathcal{X}} m(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x}) d\boldsymbol{x} \int_{\mathcal{X}} m(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x})' d\boldsymbol{x} \left\{ \int_{\mathcal{X}} \boldsymbol{u}(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x})' d\boldsymbol{x} \right\}^{-1} \right\} \\
&= \left\{ \int_{\mathcal{X}} \boldsymbol{u}(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x})' d\boldsymbol{x} \right\}^{-1} \int_{\mathcal{X}} \int_{\mathcal{X}} E\{m(\boldsymbol{x}) m(\boldsymbol{t})\} \boldsymbol{u}(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{t})' d\boldsymbol{x} d\boldsymbol{t} \left\{ \int_{\mathcal{X}} \boldsymbol{u}(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x})' d\boldsymbol{x} \right\}^{-1} \\
&= \tau^2 \left\{ \int_{\mathcal{X}} \boldsymbol{u}(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x})' d\boldsymbol{x} \right\}^{-1} \int_{\mathcal{X}} \int_{\mathcal{X}} \psi(\boldsymbol{x}, \boldsymbol{t}) \boldsymbol{u}(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{t})' d\boldsymbol{x} d\boldsymbol{t} \left\{ \int_{\mathcal{X}} \boldsymbol{u}(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{x})' d\boldsymbol{x} \right\}^{-1}.
\end{aligned}$$

Furthermore, the stochastic integral can be written as the limit of a Riemann sum, where each sum follows a normal distribution with bounded variance. Therefore, the limit is also a normally distributed random variable, which completes the proof. □

Corollary 1: If Legendre polynomials are used as the basis functions, then

$$\boldsymbol{\beta} \sim N\left(\boldsymbol{0}, \int_{\mathcal{X}} \int_{\mathcal{X}} \psi(\boldsymbol{x}, \boldsymbol{t}) \boldsymbol{u}(\boldsymbol{x}) \boldsymbol{u}(\boldsymbol{t})' d\boldsymbol{x} d\boldsymbol{t} \right).$$

It might be possible to derive analytical expressions for some of the correlation functions such as cubic correlation. However, we postpone it for future work. For the immediate purpose, we use simulations to compute the prior distribution. Choose a set of points $\boldsymbol{S} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k\}$ from $\mathcal{X}$. The points should fill-in the space $\mathcal{X}$ well to get a good approximation. Note that $\boldsymbol{S}$ is different from the experimental design $\boldsymbol{D} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ that is used for collecting the data. The size of $\boldsymbol{S}$ should be much larger than the size of $\boldsymbol{D}$, i.e., $k >> n$. Let $\boldsymbol{U}_S$ be the $k \times (N+1)$ model matrix generated from $\boldsymbol{u}(\boldsymbol{x})'\boldsymbol{\beta}$ based on the points in $\boldsymbol{S}$ and $\boldsymbol{\Psi}$ be the $k \times k$ correlation matrix for the points in $\boldsymbol{S}$ whose ijth element is $\psi(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Then, we obtain the following result.

THEOREM 2: Suppose the GP will be observed only at the points in $\boldsymbol{S}$. Then, a linear model under the prior

$$\boldsymbol{\beta} \sim N\left(\mathbf{0}, \tau^2(\boldsymbol{U}_S'\boldsymbol{U}_S)^{-1}\boldsymbol{U}_S'\boldsymbol{\Psi}\boldsymbol{U}_S(\boldsymbol{U}_S'\boldsymbol{U}_S)^{-1}\right)$$

will give the best match to the GP prior in the sense of least squares.

Proof. The least squares solution of $\boldsymbol{\beta}$ is given by $\boldsymbol{\beta} = (\boldsymbol{U}_S'\boldsymbol{U}_S)^{-1}\boldsymbol{U}_S'\boldsymbol{m}$, where $\boldsymbol{m} = (m(\boldsymbol{x}_1),\cdots, m(\boldsymbol{x}_k))'$. It is easy to see that: $E(\boldsymbol{\beta}) = \mathbf{0}$ and
$var(\boldsymbol{\beta}) = \tau^2(\boldsymbol{U}_S'\boldsymbol{U}_S)^{-1}\boldsymbol{U}_S'\boldsymbol{\Psi}\boldsymbol{U}_S(\boldsymbol{U}_S'\boldsymbol{U}_S)^{-1}$. Moreover, because $\boldsymbol{\beta}$ is a linear combination of normally distributed random variables, it also follows a normal distribution. Thus, we obtain the desired result. □

Theorem 2 is a generalization of the functionally induced priors proposed in Joseph (2006). To see this, let $\boldsymbol{S}$ be the full factorial design obtained based on the levels of the factors in the experimental design $\boldsymbol{D}$. Also expand the basis to include all of the factorial effects that can be estimated using $\boldsymbol{S}$. Now, $\boldsymbol{U}_S$ becomes a square matrix and invertible and thus, we obtain $\boldsymbol{\beta} \sim N\left(\mathbf{0}, \tau^2\boldsymbol{U}_S^{-1}\boldsymbol{\Psi}\boldsymbol{U}_S'\right)$, which is the prior distribution proposed in Joseph (2006) and Joseph and Delaney (2007). They have shown that the functionally induced priors obtained by using a product correlation function satisfy the principles of effect hierarchy and heredity (Hamada and Wu 1992). We can expect to see a similar behavior for the prior specified in Theorem 2 as well.

Now that we have a prior for $\boldsymbol{\beta}$ that approximates the GP prior of $m(\boldsymbol{x})$, we can find the posterior mean of $\boldsymbol{\beta}$ as in Proposition 1. Let $\boldsymbol{\Sigma} = (\boldsymbol{U}_S'\boldsymbol{U}_S)^{-1}\boldsymbol{U}_S'\boldsymbol{\Psi}\boldsymbol{U}_S(\boldsymbol{U}_S'\boldsymbol{U}_S)^{-1}$ and $\boldsymbol{U}_D$ be the $n \times (N+1)$ model matrix generated based on the points in the experimental design $\boldsymbol{D}$. Then, the posterior mean of $\boldsymbol{\beta}$ can be obtained as

$$\tilde{\boldsymbol{\beta}} = \tau^2\boldsymbol{\Sigma}\boldsymbol{U}_D'(\tau^2\boldsymbol{U}_D\boldsymbol{\Sigma}\boldsymbol{U}_D' + \nu^2\boldsymbol{I}_n + \frac{\sigma^2}{r}\boldsymbol{I}_n)^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{f}(\tilde{\boldsymbol{\eta}}) - \boldsymbol{F}(\tilde{\boldsymbol{\eta}})\tilde{\boldsymbol{\alpha}}). \tag{15}$$

This can be viewed as a modified ridge regression estimate obtained by minimizing the sum of squares of errors plus a quadratic penalty $\sigma^2/(r\tau^2)\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}$ instead of the usual penalty $\kappa^2\boldsymbol{\beta}'\boldsymbol{\beta}$, where κ^2 is a constant. This is the base estimate of $\boldsymbol{\beta}$ to be used in the nn-garrote method. Because a linear model is used for $m(\boldsymbol{x})$, we should update the estimate of experimental bias also. The posterior mean of $\boldsymbol{e}$ is given by

$$\tilde{\boldsymbol{e}} = \nu^2(\tau^2\boldsymbol{U}_D\boldsymbol{\Sigma}\boldsymbol{U}_D' + \nu^2\boldsymbol{I}_n + \frac{\sigma^2}{r}\boldsymbol{I}_n)^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{f}(\tilde{\boldsymbol{\eta}}) - \boldsymbol{F}(\tilde{\boldsymbol{\eta}})\tilde{\boldsymbol{\alpha}}), \tag{16}$$

which can be used as the base estimate of $\boldsymbol{e}$ in the nn-garrote method. Because we use ridge regression estimates as the base estimates, it makes sense to add the quadratic penalty terms on $\boldsymbol{\beta}$ and $\boldsymbol{e}$ to the objective function in (6). Thus the objective function becomes

$$(\bar{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{\eta}) - \boldsymbol{X}(\boldsymbol{\eta})\boldsymbol{B}\boldsymbol{\theta})'(\bar{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{\eta}) - \boldsymbol{X}(\boldsymbol{\eta})\boldsymbol{B}\boldsymbol{\theta}) + \boldsymbol{\theta}'\boldsymbol{B}\boldsymbol{R}\boldsymbol{B}\boldsymbol{\theta}, \tag{17}$$

where $\boldsymbol{X}(\boldsymbol{\eta}) = [\boldsymbol{F}(\boldsymbol{\eta}), \boldsymbol{U}, \boldsymbol{I}_n]$, $\boldsymbol{B}$ is a diagonal matrix with entries $\{\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\beta}_0, \cdots, \tilde{e}_n\}$, and $\boldsymbol{R}$ is a diagonal block matrix with blocks $\mathbf{0}_2$, $\sigma^2/(r\tau^2)\boldsymbol{\Sigma}^{-1}$, and $\sigma^2/(r\nu^2)\boldsymbol{I}_n$. With the addition of the quadratic penalty terms, the variable selection procedure can now be viewed as a version of the elastic net procedure (Zou and Hastie 2005).

To summarize the procedure, first we obtain $\tilde{\boldsymbol{\eta}}$ from (11) and $\tilde{\boldsymbol{\alpha}}$, $\tilde{\tau}^2$, $\tilde{\nu}^2$, and $\tilde{\boldsymbol{c}}$ from (12) and (13). Then, we construct the variance-covariance matrix $\boldsymbol{\Sigma}$ using the result in Theorem 2 and use (15) and (16) to obtain $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{e}}$. Then, we minimize (17) with respect to $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ subject to the nn-garrote constraints for a sequence of values of M starting with 0. The best value of M is chosen by maximizing the predictive ability of the model. One way to do this is to find M by minimizing the mean squared leave-one-out cross validation ($MSCV$) error computed as

$$MSCV(M) = \frac{1}{n}\sum_{i=1}^{n}\{\bar{y}_{i\cdot} - \hat{y}_{(i)}\}^2,$$

where $\hat{y}_{i(i)}$ is the predicted value of y at $\boldsymbol{x} = \boldsymbol{x}_i$ without using the data $\bar{y}_{i\cdot}$. It is important to note that the e_i's are present only in the experiment and therefore should not be included in obtaining the predicted value $\hat{y}_{i(i)}$. Let M^* be the value that minimizes $MSCV(M)$ and $\hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\theta}}$ be the estimates of the parameters at $M = M^*$. Then, the prediction model is given by

$$\hat{y}(\boldsymbol{x}) = f(\boldsymbol{x}; \hat{\boldsymbol{\eta}}) + \hat{\theta}_0\tilde{\alpha}_0 + \hat{\theta}_1\tilde{\alpha}_1(f(\boldsymbol{x}; \hat{\boldsymbol{\eta}}) - \bar{f}(\hat{\boldsymbol{\eta}})) + \sum_{k=0}^{N}\hat{\theta}_{k+2}\tilde{\beta}_k u_k(\boldsymbol{x}). \tag{18}$$

Again note that the prediction model does not include the experimental bias terms.

2.3.3 Standard Errors

The standard errors for the parameter estimates can be obtained using a local quadratic approximation (Fan and Li 2001, Xiong 2010). Consider an equivalent formulation

of the optimization problem:

$$\min_{\boldsymbol{\eta},\boldsymbol{\theta}} (\bar{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{\eta}) - \boldsymbol{X}(\boldsymbol{\eta})\boldsymbol{B}\boldsymbol{\theta})'(\bar{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{\eta}) - \boldsymbol{X}(\boldsymbol{\eta})\boldsymbol{B}\boldsymbol{\theta}) + \boldsymbol{\theta}'\boldsymbol{B}\boldsymbol{R}\boldsymbol{B}\boldsymbol{\theta} + \xi\boldsymbol{\theta}'\mathbf{1}_{N+n+1}$$

subject to the nonnegativity constraints on $\boldsymbol{\theta}$, where ξ is the Lagrangian multiplier. Suppose we replace the linear constraint $\sum_{i=0}^{N+n+1}\theta_i$ with a quadratic constraint $\sum_{i=0}^{N+n+1}\theta_i^2/\hat{\theta}_i$ and $\boldsymbol{X}(\boldsymbol{\eta})$ with $\boldsymbol{X}(\hat{\boldsymbol{\eta}})$, then the solution will be approximately the same as before. Because some $\hat{\theta}_i$'s can be 0, we should only use the terms in the model with $\hat{\theta}_i > 0$. However, for simplicity, we will not introduce any new notations for $\boldsymbol{X}$, $\boldsymbol{B}$, and $\boldsymbol{\theta}$ after removing the terms. Thus, the optimization problem can be written as minimizing

$$L = (\bar{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{\eta}) - \boldsymbol{X}(\hat{\boldsymbol{\eta}})\boldsymbol{B}\boldsymbol{\theta})'(\bar{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{\eta}) - \boldsymbol{X}(\hat{\boldsymbol{\eta}})\boldsymbol{B}\boldsymbol{\theta}) + \boldsymbol{\theta}'\boldsymbol{B}\boldsymbol{R}\boldsymbol{B}\boldsymbol{\theta} + \xi\boldsymbol{\theta}'\boldsymbol{S}^{-1}\boldsymbol{\theta},$$

where $\boldsymbol{S}$ is a diagonal matrix with entries $\hat{\boldsymbol{\theta}}$ (for which $\hat{\theta}_i > 0$). First, using a Taylor series approximation: $f(\boldsymbol{x};\boldsymbol{\eta}) \approx f(\boldsymbol{x};\hat{\boldsymbol{\eta}}) + \nabla f(\boldsymbol{x};\hat{\boldsymbol{\eta}})(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})$, where $\nabla f(\boldsymbol{x};\boldsymbol{\eta})$ is the $1 \times q$ gradient vector of $f(\boldsymbol{x};\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$. Now, differentiating L with respect to $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$, and equating to 0, we obtain

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\eta}} &= -2\nabla\boldsymbol{f}(\hat{\boldsymbol{\eta}})'\{\bar{\boldsymbol{y}} - \boldsymbol{f}(\hat{\boldsymbol{\eta}}) - \nabla\boldsymbol{f}(\hat{\boldsymbol{\eta}})(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}) - \boldsymbol{X}(\hat{\boldsymbol{\eta}})\boldsymbol{B}\boldsymbol{\theta}\} = 0, \\
\frac{\partial L}{\partial \boldsymbol{\theta}} &= -2\boldsymbol{B}\boldsymbol{X}(\hat{\boldsymbol{\eta}})'\{\bar{\boldsymbol{y}} - \boldsymbol{f}(\hat{\boldsymbol{\eta}}) - \nabla\boldsymbol{f}(\hat{\boldsymbol{\eta}})(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}) - \boldsymbol{X}(\hat{\boldsymbol{\eta}})\boldsymbol{B}\boldsymbol{\theta}\} + 2(\xi\boldsymbol{S}^{-1} + \boldsymbol{B}\boldsymbol{R}\boldsymbol{B})\boldsymbol{\theta} = 0,
\end{aligned}$$

where $\nabla\boldsymbol{f}(\hat{\boldsymbol{\eta}})$ is the $n \times q$ matrix with ith row $\nabla f(\boldsymbol{x}_i;\hat{\boldsymbol{\eta}})$. The solution to these two system of equations is given by

$$\begin{pmatrix} \boldsymbol{\eta} - \hat{\boldsymbol{\eta}} \\ \boldsymbol{\theta} \end{pmatrix} = \boldsymbol{H}_\xi^{-1} \begin{pmatrix} \nabla\boldsymbol{f}(\hat{\boldsymbol{\eta}})'\{\bar{\boldsymbol{y}} - \boldsymbol{f}(\hat{\boldsymbol{\eta}})\} \\ \boldsymbol{B}\boldsymbol{X}(\hat{\boldsymbol{\eta}})'\{\bar{\boldsymbol{y}} - \boldsymbol{f}(\hat{\boldsymbol{\eta}})\} \end{pmatrix},$$

where

$$\boldsymbol{H}_\xi = \begin{pmatrix} \nabla\boldsymbol{f}(\hat{\boldsymbol{\eta}})'\nabla\boldsymbol{f}(\hat{\boldsymbol{\eta}}) & \nabla\boldsymbol{f}(\hat{\boldsymbol{\eta}})'\boldsymbol{X}(\hat{\boldsymbol{\eta}})\boldsymbol{B} \\ \boldsymbol{B}\boldsymbol{X}(\hat{\boldsymbol{\eta}})'\nabla\boldsymbol{f}(\hat{\boldsymbol{\eta}}) & \boldsymbol{B}\boldsymbol{X}(\hat{\boldsymbol{\eta}})'\boldsymbol{X}(\hat{\boldsymbol{\eta}})\boldsymbol{B} + \xi\boldsymbol{S}^{-1} + \boldsymbol{B}\boldsymbol{R}\boldsymbol{B} \end{pmatrix}.$$

Now ξ can be chosen to get the solution close to the solution of the original problem. This can be done by minimizing a weighted sum of squares of the elements in $\frac{\partial L}{\partial \boldsymbol{\theta}}$ given by $\frac{\partial L}{\partial \boldsymbol{\theta}}'\boldsymbol{S}\frac{\partial L}{\partial \boldsymbol{\theta}}$ evaluated at $(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}})$ with respect to ξ. We obtain

$$\xi^* = \frac{\hat{\boldsymbol{\theta}}'[\boldsymbol{B}\boldsymbol{X}(\hat{\boldsymbol{\eta}})'\{\bar{\boldsymbol{y}} - \boldsymbol{f}(\hat{\boldsymbol{\eta}}) - \boldsymbol{X}(\hat{\boldsymbol{\eta}})\boldsymbol{B}\hat{\boldsymbol{\theta}}\} - \boldsymbol{B}\boldsymbol{R}\boldsymbol{B}\hat{\boldsymbol{\theta}}]}{M^*}.$$

Thus, an estimate of the variance of $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\theta}}$ is given by

$$\widehat{var}\begin{pmatrix}\hat{\boldsymbol{\eta}}\\ \hat{\boldsymbol{\theta}}\end{pmatrix} = \boldsymbol{H}_{\xi^*}^{-1}\widehat{var}\begin{pmatrix}\nabla \boldsymbol{f}(\hat{\boldsymbol{\eta}})'\{\bar{\boldsymbol{y}}-\boldsymbol{f}(\hat{\boldsymbol{\eta}})\}\\ \boldsymbol{B}\boldsymbol{X}(\hat{\boldsymbol{\eta}})'\{\bar{\boldsymbol{y}}-\boldsymbol{f}(\hat{\boldsymbol{\eta}})\}\end{pmatrix}\boldsymbol{H}_{\xi^*}^{-1} = \frac{\sigma^2}{r}\boldsymbol{H}_{\xi^*}^{-1}\boldsymbol{V}\boldsymbol{H}_{\xi^*}^{-1}, \tag{19}$$

where

$$\boldsymbol{V} = \begin{pmatrix} \nabla \boldsymbol{f}(\hat{\boldsymbol{\eta}})'\nabla \boldsymbol{f}(\hat{\boldsymbol{\eta}}) & \nabla \boldsymbol{f}(\hat{\boldsymbol{\eta}})'\boldsymbol{X}(\hat{\boldsymbol{\eta}})\boldsymbol{B} \\ \boldsymbol{B}\boldsymbol{X}(\hat{\boldsymbol{\eta}})'\nabla \boldsymbol{f}(\hat{\boldsymbol{\eta}}) & \boldsymbol{B}\boldsymbol{X}(\hat{\boldsymbol{\eta}})'\boldsymbol{X}(\hat{\boldsymbol{\eta}})\boldsymbol{B} \end{pmatrix}.$$

Note that if $\boldsymbol{\theta}$ were not in the model, then this formula reduces to the well-known formula for the variance of nonlinear least squares estimate of $\boldsymbol{\eta}$ (Bates and Watts 1988). Similarly, if $\boldsymbol{\eta}$ were not in the model, then the formula reduces to the formula for nn-garrote parameter estimates given in Xiong (2010).

Uncertainty in the prediction can be quantified using

$$var\{\hat{y}(\boldsymbol{x})\} \approx \frac{\sigma^2}{r}\boldsymbol{w}(\boldsymbol{x})'\boldsymbol{H}_{\xi^*}^{-1}\boldsymbol{V}\boldsymbol{H}_{\xi^*}^{-1}\boldsymbol{w}(x), \tag{20}$$

where $\boldsymbol{w}(\boldsymbol{x}) = (\nabla f(\boldsymbol{x};\hat{\boldsymbol{\eta}})', \boldsymbol{v}(\boldsymbol{x})'\boldsymbol{B})'$ and $\boldsymbol{v}(\boldsymbol{x}) = (1, f(\boldsymbol{x};\hat{\boldsymbol{\eta}})-\bar{f}(\hat{\boldsymbol{\eta}}), u_1(\boldsymbol{x}), \cdots, u_N(\boldsymbol{x}))'$. Here we have assumed that the physical model is known. Many times the physical model is a complex computer code and therefore, it can only be observed at locations specified in a computer experiment. A metamodel such as a kriging model can be fitted using these computer experiment data to approximate the physical model, but it introduces uncertainties at the unobserved locations. We can transmit these uncertainties into the prediction variance to get a more realistic estimate. Let $g(\boldsymbol{x})$ be the prediction variance of the metamodel at $\boldsymbol{x}$. For example, suppose a simple kriging model with a stationary covariance function $\gamma^2 R(.)$ is used. Then $g(\boldsymbol{x}) = \gamma^2(1-\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}))$, where $\boldsymbol{r}(\boldsymbol{x})$ is a vector with ith element $R(\boldsymbol{x}-\boldsymbol{x}_i)$ and $\boldsymbol{R}$ is a matrix with ijth element $R(\boldsymbol{x}_i-\boldsymbol{x}_j)$. Thus,

$$\begin{aligned} var\{\hat{y}(\boldsymbol{x})\} &= E\{var[\hat{y}(\boldsymbol{x})|f(\boldsymbol{x};\hat{\boldsymbol{\eta}})]\} + var\{E[\hat{y}(\boldsymbol{x})|f(\boldsymbol{x};\hat{\boldsymbol{\eta}})]\} \\ &\approx \tfrac{\sigma^2}{r}\boldsymbol{w}(\boldsymbol{x})'\boldsymbol{H}_{\xi^*}^{-1}\boldsymbol{V}\boldsymbol{H}_{\xi^*}^{-1}\boldsymbol{w}(x) + (1+\hat{\theta}_1\hat{\alpha}_1)^2 g(\boldsymbol{x}), \end{aligned} \tag{21}$$

where the first term is again an approximation obtained by plugging-in the metamodel for $f(\boldsymbol{x};\hat{\boldsymbol{\eta}})$ in (20).

2.4 Examples

Two real examples are considered in this section. The first one on laser milling has two variables and no calibration parameters and the second one on spot welding has three variables and one calibration parameter.

2.4.1 Laser Milling Example

Dobrev, Dimov, and Thomas (2006) proposed a simulation model to understand the effect of laser fluence on the crater depth in a laser milling process. They also conducted a physical experiment using two different tooling materials: *Aluminium 6082* and *Stainless Steel 316* . Thus in this example, there are two variables, laser fluence (x_1) and material type (x_2). The simulation model, which uses a forward finite-difference method to solve a heat equation, was run at 9 laser fluence values for each of the materials. The physical experiment was conducted at 6 levels of the laser fluence under each of the material types with three replicates. Dorev et al. (2006) did not provide details on how the replicates were obtained, but as explained in Section 3, we will be able to fit the models without worrying about the type of randomization in their experiment.

The first step is to approximate the complex simulation model using an interpolating model such as kriging (Sacks et al. 1989). We used a Gaussian correlation function for the laser fluence. Because the material type is a qualitative factor, an isotropic correlation function was employed (Joseph and Delaney 2007, Qian, Wu, and Wu 2008). The crater depth is a nonnegative variable and it goes to 0 as the laser fluence is reduced to 0. Directly fitting an ordinary kriging model to the simulation data can violate these physical restrictions. Therefore, the following kriging model is fitted:

$$f(\boldsymbol{x}) = x_1^3(0.567 + 0.262x_2 + Z(\boldsymbol{x})),$$

where $Z(\boldsymbol{x}) \sim GP(0, \gamma^2 R)$ with

$$R(\boldsymbol{x}_j - \boldsymbol{x}_k) = \exp\{-d_1(x_{1j} - x_{1k})^2 - d_2 \mathrm{I}_{[x_{2j} - x_{2k} \neq 0]}\}.$$

Here the model $x_1^3(0.567 + 0.262x_2)$ was estimated using nonlinear regression. These parameters could also be estimated simultaneously along with the correlation parameters, but in our experience, the simplicity of the foregoing procedure outweigh the possible improvement in the estimation. The d_1 and d_2 are estimated as 14.096 and 1.530, respectively.

Now we are ready to apply the model calibration procedure. The fluence variable is transformed using the Legendre orthogonal polynomials up to degree 5 (because of the 6 levels). The material type (x_2) is coded as -1 and 1 to represent aluminium

and stainless steel. Using (11) and (12), we obtain: $\tilde{\tau}^2 = 0.00001$, $\tilde{\nu}^2 = 0.8123$, $c_1 = 0.6789$, $c_2 = 0.0126$, $\tilde{\alpha}_0 = -1.555$ and $\tilde{\alpha}_1 = -0.314$. The prior variance-covariance matrix of $\boldsymbol{\beta}$ is obtained using the result of Theorem 2 based on 200 randomly selected points from $\mathcal{X} = [-1, 1] \times \{-1, 1\}$. Then, we use equations (15) and (16) to obtain the base estimates of $\boldsymbol{\beta}$ and $\boldsymbol{e}$.

The nn-garrote was then applied by varying M at the intervals of 0.01 and the value that minimizes $MSCV$ is obtained as $M^* = 4.28$. At this value of M, there are five non-zero θ_i's corresponding to $\tilde{\alpha}_0$, $\tilde{\alpha}_1$, $\tilde{e}_3$, $\tilde{e}_4$, and $\tilde{e}_{10}$. Thus, the engineering model requires only a location and scale adjustment. The remaining discrepancies can be attributed to the experimental biases. Thus, the prediction model is given by

$$\hat{y}(\boldsymbol{x}) - f(\boldsymbol{x}) = -1.397 - 0.301(f(\boldsymbol{x}) - 6.816).$$

For comparison, we also consider the model

$$y - f(\boldsymbol{x}) = \alpha_0 + m(\boldsymbol{x}) + \epsilon, \;\; m(\boldsymbol{x}) \sim GP(0, \tau^2\psi), \tag{22}$$

as in Bayarri et al. (2007), which is a refined version of the model in Kennedy and O'Hagan (2001). We call this the full adjustment model, because all the discrepancies are used for adjusting the engineering model. Figure 3 shows the predictions from all the three models, viz., engineering model, minimal adjustment model, and the full adjustment model. We can see that the full adjustment model gives the best fit to the data. However, a better fit does not necessarily indicate a better model. The data may contain some experimental biases and fitting to such data will not make a good prediction model. The minimal adjustment model, on the other hand, makes only small changes to the engineering model but gives a reasonable fit to the data. We can see that the minimal adjustment model preserves the shape of the engineering model but corrects for its discrepancy. Thus, instead of over fitting the data, it respects the information contained in the engineering model and brings it closer to the data. This is expected to make a good prediction model.

In the foregoing analysis we had tacitly assumed that the 12 settings (6 laser settings under each of the two material types) were done independent of each other. To reduce experimental cost, it is common to apply restrictions on randomization such as fixing the laser setting at a level and completing the experiments on the two

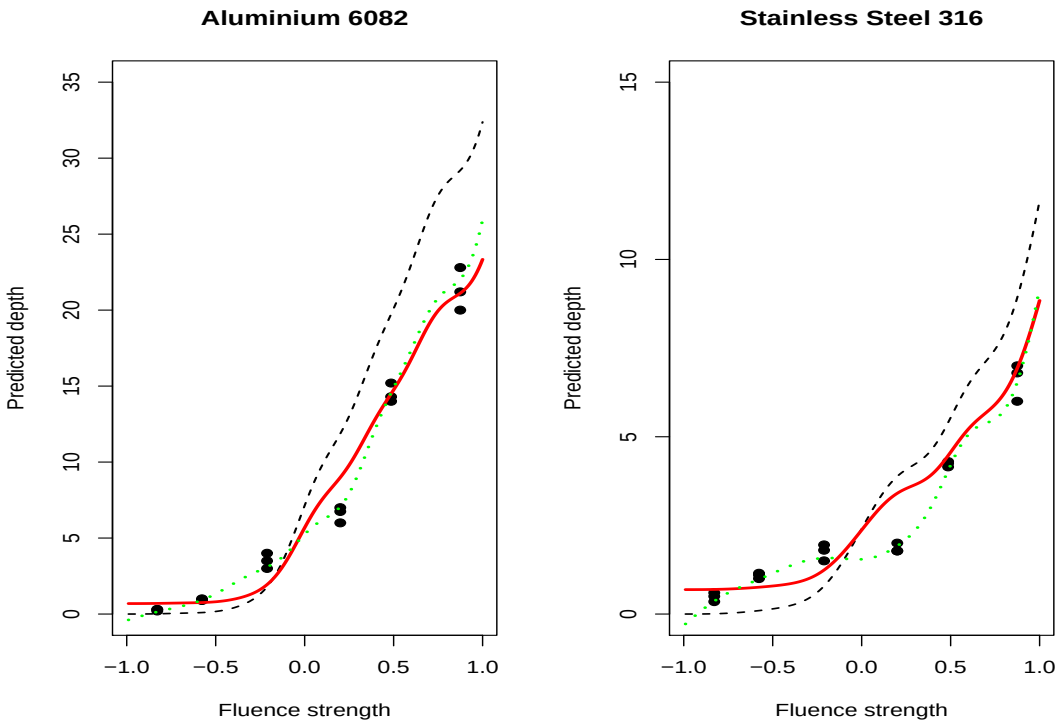


Figure 3: Physical model (Dashed), minimal adjustment model (Solid) and full adjustment model (Dotted) in the laser milling example.

material types. This introduces a positive correlation among the e_i's (within each laser setting). Recognizing such model structures can help estimate the experimental bias terms more accurately. We are not able to use such information here because the details on how the experiment was conducted were not available in Dorev et al. (2006).

2.4.2 Spot Welding Example

Consider the spot welding example discussed in Higdon et. al. (2004) and Bayarri et. al. (2007). There are three controllable variables in this experiment: load (x_1), current (x_2), and gage (x_3). In addition, the engineering model also contains an unknown calibration parameter (η). The ultimate goal of this experiment is to optimize the nugget diameter of the weld (y).

The physical model to simulate the spot welding process is a computationally expensive finite element model (FEM). Therefore, an easy-to-evaluate approximate model is necessary to perform the model calibration. For this purpose, a computer experiment is carried out using the FEM according to a 52-run maximin Latin hypercube design. After removing 17 runs that did not produce clear outputs, the remaining 35 runs (see Table 3 of Bayarri et al. 2007) were used for fitting an ordinary kriging model, which will be used in place of the FEM. The physical experiment has 12 runs with 10 replicates per run (see Table 4 of Bayarri et al. 2007). We proceed as before for fitting the minimal adjustment model. The computation here is much more complex than the laser milling example due to the presence of the calibration parameter, which needs to be estimated using nonlinear least squares at each value of M. The MSCV is minimized at $M^* = 6.01$. The fitted model at M^* contains four model bias terms and five experimental bias terms (e_4, e_5, e_8, e_{10}, and e_{12}). The prediction model is given by

$$\begin{aligned}\hat{y}(\boldsymbol{x}) - f(\boldsymbol{x}; 2.049) = \; & -0.3667(f(\boldsymbol{x}; 2.049) - \bar{f}(2.049)) \\ & +0.652u_3 + 0.061u_1u_2 + 0.035u_1u_3 + 0.116u_2u_3,\end{aligned}$$

where u_1, u_2 and u_3 represent the scaled first order Legendre orthogonal polynomials of x_1, x_2 and x_3, respectively. The estimation of the unknown calibration parameter (η) is equal to 2.049. Its 95% confidence interval can be calculated using (19) as: $(1.917, 2.181)$.

In Figure 4, we plot the predictions of the nugget diameter from three different models including the engineering model (with η fixed at its least squares estimate $\tilde{\eta} = 2.1$), the proposed minimal adjustment model, and the full adjustment model. We can see that the minimal adjustment and full adjustment models fit the data much better than the engineering model. The full adjustment model gives good fit, but it looks quite different from the engineering model. On the other hand, the minimal adjustment model brought the engineering model closer to the data with only a slight change in the shape that is not even noticeable. The major discrepancy of the minimal adjustment model compared to the data can be seen at the right top panel of Figure 4. However, this discrepancy is identified as the experimental bias e_8. Note that if we were to classify this discrepancy as a model bias term, then a major change in the shape of the engineering model would result, which is unlikely to be true.

The 95% prediction intervals calculated using (21) are shown in Figure 5. They look narrower compared to the prediction intervals given in Bayarri et al. (2007). On the one hand, the approach in Bayarri et al. is Bayesian which can more effectively incorporate all of the uncertainties and thus the prediction variance obtained in (21) could be an underestimate of the actual prediction variance. On the other hand, the prediction variance in Bayearri et al. could be an overestimate because of not removing the experimental biases from the data.

2.5 Simulations

Although the minimal adjustment procedure seems to be effective in the two real examples discussed in the previous section, the results are not conclusive because we do not know the true models. Here, we generate data from some known models and check the performance of the proposed procedure.

Consider the following simple example similar to the one discussed in Section 1. Let the data are generated from $y_{ij} = 10(e^{0.01x_i} - 1) + k\sin(0.07\pi x_i) + \epsilon_{ij}$, for $i = 1, 2, \ldots, 7$, and $j = 1, \ldots, 10$, where $x_i = \{0, 10, 20, 30, 40, 50, 60\}$ and $\epsilon_{ij} \sim N(0, \sigma^2)$. Suppose that the physical model is a first-order linear approximation of the true model given by $f(x_i) = 0.1x_i$. In the Section 1, we introduced a shift of 3 units at the middle observation. Here we perform a similar simulation by randomizing the

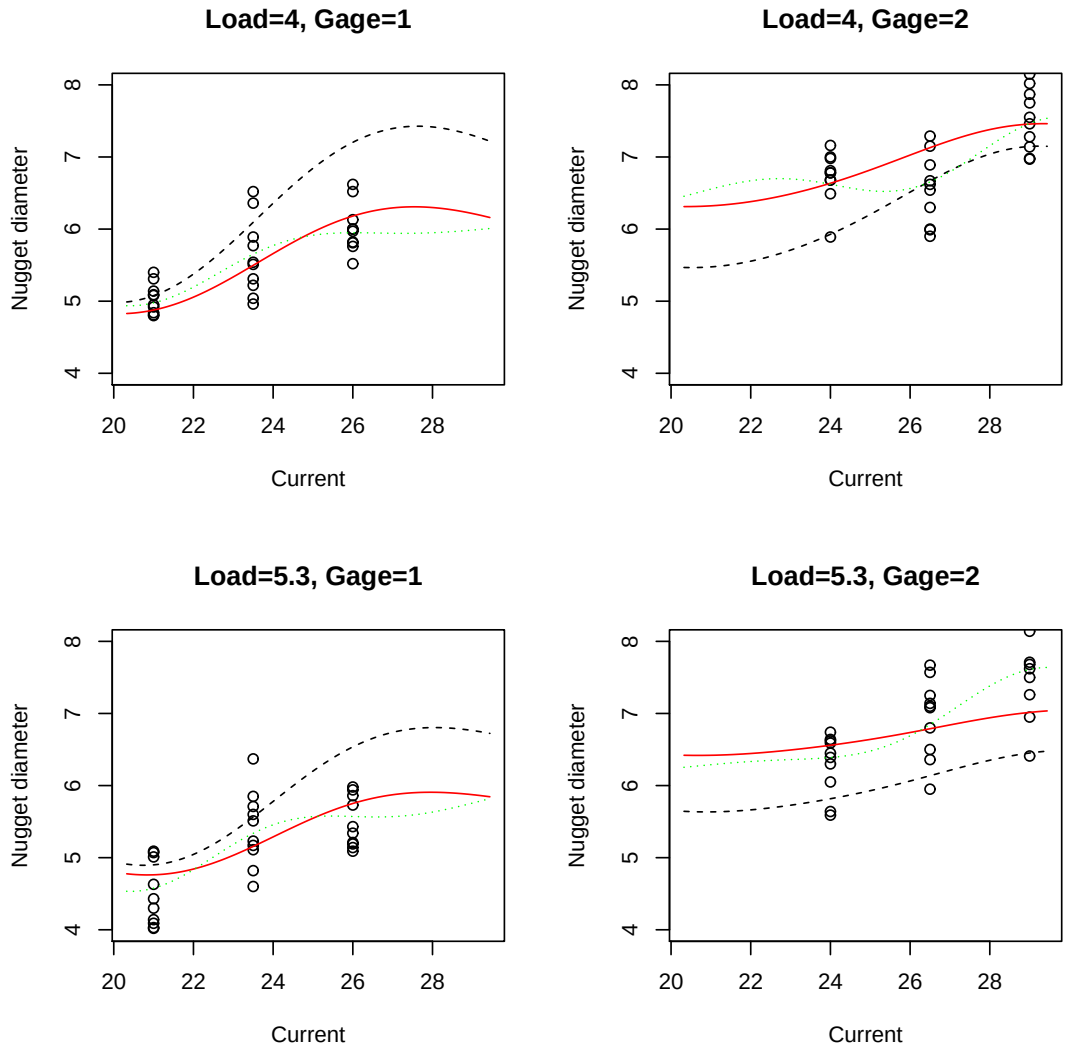


Figure 4: Physical model with $\eta = 2.10$ (Dashed), minimal adjustment model (Solid), and full adjustment model (Dotted) in the spot welding example.

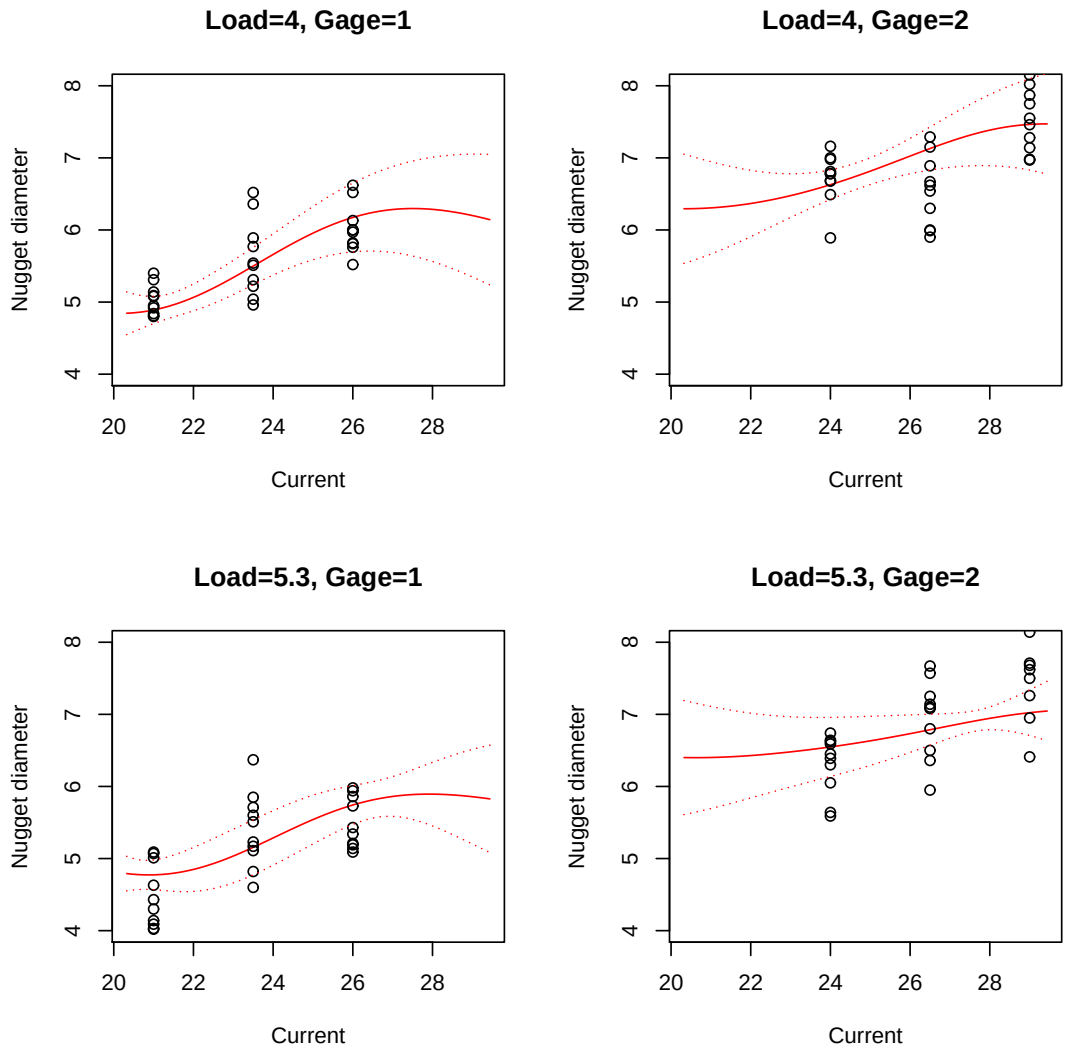


Figure 5: Minimal adjustment model (Solid) and its 95% confidence interval (Dotted) in the spot welding example.

magnitude and location of the shift. Moreover, we vary k between 0 and 1 so that we can learn more about the effect of model complexity.

First consider the simulations by randomizing the magnitude of the shift at a fixed location. We generate the shift from $U(-4.4, 0)$ when $k = 0$ (and shift from $U(-5.6, 0)$ when $k = 1$) and add it to the observations at the middle ($i = 4$). The values -4.4 for $k = 0$ and -5.6 for $k = 1$ were chosen so that the average experimental bias is the same as the maximum model bias. Then, we fit both minimal adjustment model and full adjustment model in (22). The maximum absolute error (MAE) of the fitted model is adopted as the performance evaluation criteria which is calculated as

$$MAE = \max_i \{\hat{y}(t_i) - 10(e^{0.01t_i} - 1)\}^2,$$

where $t_1, \ldots, t_{601}$ are equally spaced values from 0 to 60. The MAEs for 500 simulations with $\sigma^2 = 1$ are plotted in Figures 6 (a) and (b). We can see that minimal adjustment performs better than full adjustment (88.0% of the cases when $k = 0$ and 82.8% of the cases when $k = 1$.) Note that the performance of minimal adjustment decreases as the model complexity increases. In the figures we have used two plotting symbols based on the size of the shift. As expected, the performance of minimal adjustment is better with larger shift size.

Now consider the simulations by randomizing the location of the shift instead of the magnitude of the shift. In each of the 500 simulations, the location of the shift is randomly chosen from $x_i = \{0, 10, 20, 30, 40, 50, 60\}$. The MAEs for $k = 0$ and $k = 1$ are plotted in Figures 6 (c) and 6 (d). Here the minimal adjustment performs better than the full adjustment in 87.8% and 77.6% of the cases corresponding to $k = 0$ and $k = 1$.

The simulations are repeated for $\sigma^2 = 0.1$ and we found that the performance of minimal adjustment is even better (details are not shown here). This is because it is easier to identify and remove the shift with a smaller σ^2, which makes the minimal adjustment more effective.

We now consider a simulation model that contains a calibration parameter: $y_{ij} = 10(e^{0.1\eta x_i} - 1) + k\sin(0.07\pi x_i) + \epsilon_{ij}$, for $i = 1, 2, \ldots, 7$, and $j = 1, \ldots, 10$, where $\epsilon_{ij} \sim N(0, 1)$. The data are generated by assuming $\eta = 0.1$. The physical model in this study is taken as $f(x; \eta) = \eta x$. As before, we performed the simulations by

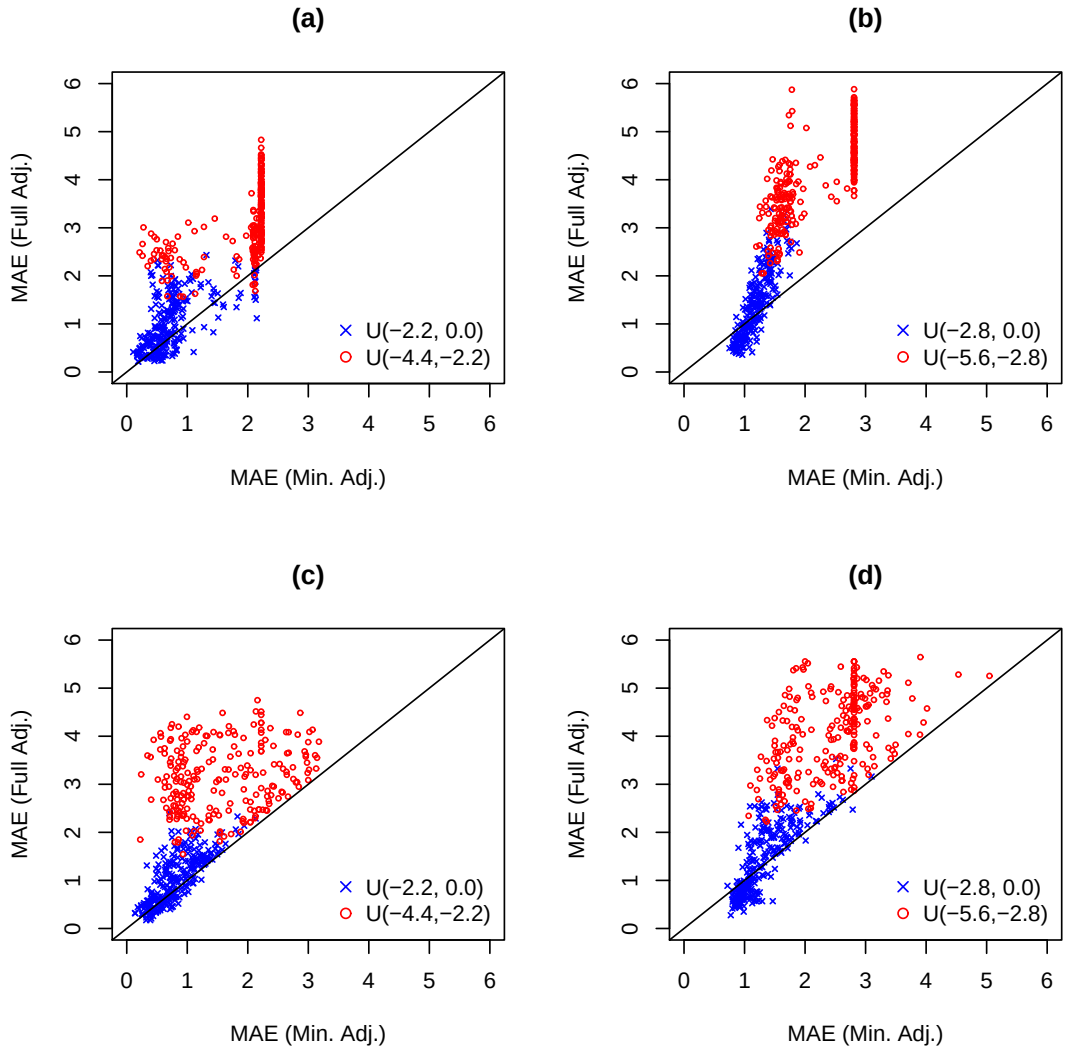


Figure 6: Comparison between minimal adjustment and full adjustment models. Fixed shift location with: (a) $k = 0$, (b) $k = 1$; Random shift location with: (c) $k = 0$, and (d) $k = 1$.

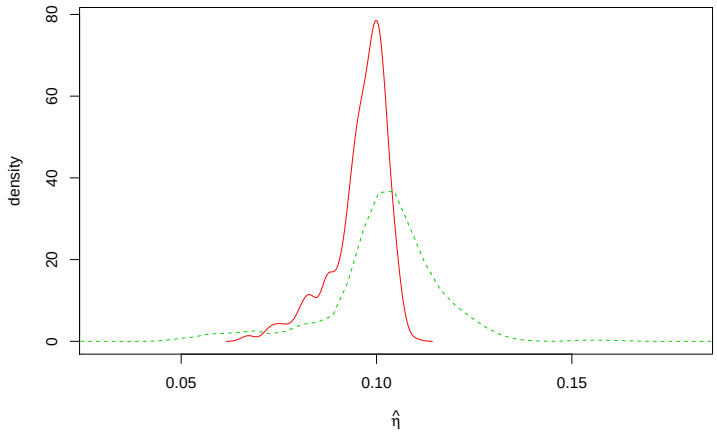


Figure 7: Density plot of calibration parameter estimated between minimal adjustment model (Solid) and full adjustment model (Dashed).

varying the magnitude and location of the shift as well as k. The minimal adjustment is found to perform better than the full adjustment model in (a) 91.4% , (b) 77.0%, (c) 90.4%, and (d) 78.2% of the cases for the simulation settings shown in Figure 6. Figure 7 shows the density plot of $\hat{\eta}$ from the two methods for the the random location case with $k = 1$. We can see that the estimates from the minimal adjustment are more concentrated around the true value than those from the full adjustment.

In summary, minimal adjustment method seems to work much better than the full adjustment procedure especially when large experimental biases are present in the data. However, the minimal adjustment procedure can also make mistakes in the sense that the the procedure may identify wrong model bias terms and experimental bias terms. As alluded to before, separating model and experimental biases is a difficult problem and it is not easy to develop a method that will work all the time irrespective of the situations.

2.6 Conclusions

The existing approach to calibration in the presence of model uncertainties is to fit a Gaussian process (GP) model to capture the bias in the physical model. The GP model is quite flexible and therefore,it can easily adapt to different data situations. This flexibility can be harmful when there are systematic errors in the experiments.

These errors are confounded with the model bias terms and therefore, the fitted GP model can become quite complex and different from the true model. Here we introduced a minimal adjustment procedure for calibration, which brings the physical model closer to the data by making minimal changes to it. This is achieved by a simultaneous variable selection of the model and experimental bias terms. Our procedure starts with a Gaussian process model and then approximates it using a linear regression model. The variable selection using nonnegative garrote is then applied to select the important variables in the linear regression model along with the experimental bias terms. In layman language, this can be viewed as cutting the Gaussian process model into several pieces and selecting only the pieces that are really needed for model adjustments.

Two real examples and simulations are presented to demonstrate the advantages of using the minimal adjustment procedure for model calibration. The minimal adjustment procedure respects the physical model in a better way than the GP model adjustment by acknowledging the possibility of systematic errors in the experiment. Thus, the valuable information in the physical model will not be lost due to a few experimental errors. The minimal adjustment procedure is not a mistake-free method. It can also confuse between model and experimental bias terms. However, on an average it is found to perform much better than the GP model adjustment.

Our approach is frequentist although it has a Bayesian flavor due to the use of a prior variance-covariance matrix for obtaining the ridge regression estimates, which are used as base estimates for the parameters in the nonnegative garrote. A disadvantage of the frequentist approach is that it does not account for various parameter uncertainties as in the Bayesian approach. Although this has little effect on the model predictions, the confidence intervals around the predictions can be narrower. However, this is a minor problem in the context of model calibration. As pointed out by Kennedy and O'Hagan (2001), it is much more important to account for various sources of uncertainties than to account for all parameter uncertainties. See also Bayarri et al. (2007). The incorporation of the source of experimental errors in the model calibration will improve the prediction and confidence intervals much more than trying to capture all of the parameter uncertainties.

In our approach we have assumed no prior knowledge about the experimental

bias. However, in many cases, the experimenter might know when and how they have happened. Consider a situation where the experimenter knows a bias has occurred at some setting but does not know about its magnitude (of course, if the magnitude is also known, then it can be easily removed from the data.) This knowledge can be incorporated into the procedure by removing the corresponding θ_i from the linear constraint in the nn-garrote method. Thus, no penalty will be given to the setting where a bias is known to have occurred and therefore it will always be selected in the variable selection procedure.

2.7 References

Bates, D. M., and Watts, D. G. (1988), *Nonlinear Regression Analysis and Its Applications*, New York: Wiley.

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C. H., Tu, J. (2007), "A Framework for Validation of Computer Models," *Technometrics*, 49, 138–154.

Breiman, L. (1995). "Better Subset Regression Using the Nonnegative Garrote". *Technometrics*, 37, 373–384.

Cramér, H. and Leadbetter, M. R. (1967), *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*, New York: Wiley.

Deng, X., Joseph, V. R., Mai, W., Wang, Z. L., and Wu, C. F. J. (2009), "A Statistical Approach to Quantifying Elastic Deformation of Nanomaterials," *Proceedings of the National Academy of Sciences*, 106, 11845–11850.

Dobrev, T., Dimov, S. S., and Thomas, A. J. (2006), "Laser Milling: Modelling Crater and Surface Formation," *Proceedings of the Institution of Mechanical Engineers, Part C: J. Mechanical Engineering Science*, 220, 1685–1696.

Efron, B., Johnstone, I., Hastie, T., and Tibshirani, R. (2004),"Least Angle Regression (with discussion)," *The Annals of Statistics*, 32, 407–499.

Fan, J. and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96,

1348-1360.

Hamada, M. and Wu, C. F. J. (1992), "Analysis of designed experiments with complex aliasing," *Journal of Quality Technology*, 24, 130-137.

Han, G., Santner, T. J., and Rawlinson, J. J. (2009), "Simultaneous Determination of Tuning and Calibration Parameters for Computer Experiments," *Technometrics*, 51, 464-474.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, New York: Springer.

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004), "Combining Field Data and Computer Simulations for Calibration and Prediction," *SIAM Journal of Scientific Computing*, 26, 448–466.

Hildebrand, F. B. (1987), *Introduction to Numerical Analysis*, New York: Dover Publications.

Joseph, V. R. (2006). "A Bayesian Approach to the Design and Analysis of Fractionated Experiments," *Technometrics* 48, 219–229.

Joseph, V. R. and Delaney, J. D. (2007). "Functionally Induced Priors for the Analysis of Experiments," *Technometrics* 49, 1–11.

Joseph, V. R. and Melkote, S. N. (2009), "Statistical Adjustments to Engineering Models," *Journal of Quality Technology*, 41, 362–375.

Kennedy, M. C. and O'Hagan, A. (2001), "Bayesian Calibration of Computer Models (with discussion)," *Journal of Royal Statistical Society - Series B*, 63, 425–464.

Liu, F., Bayarri, M. J., and Berger, J. O. (2009), "Modularization in Bayesian Analysis, with Emphasis on Analysis of Computer Models," *Technometrics* 4, 119–150.

Qian, Z. and Wu, C. F. J. (2005), "Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments," *Technometrics*, 50, 192–204.

Qian, Z., Wu, H. and Wu, C. F. J. (2008), "Gaussian Process Models for Computer Experiments with Qualitative and Quantitative Factors," *Technometrics*, 50, 383–396.

Reese, C. S., Wilson, A. G., Hamada, M., Martz, H. F., and Ryan, K. J. (2004), "Integrated Analysis of Computer and Physical Experiments," *Technometrics*, 46, 153–164.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, New York: Springer.

Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistics Society Series B*, 58, 267–288.

Wang, S., Chen, W., and Tsui, K.-L. (2009), "Bayesian Validation of Computer Models," *Technometrics*, 51, 439–451.

Xiong, S. (2010), "Some Notes on the Nonnegative Garrote," *Technometrics*, 52, 349-361.

Xiong, Y., Chen, W., Tsui, K.-L., and Apley, D. (2009), "A better understanding of model updating strategies in validating engineering models," *Computational Methods in Applied Mechanical Engineering*, 198, 1327–1337.

Yuan, M. and Lin, Y. (2007), "On the the nonnegative garrote estimator," *Journal of the Royal Statistics Society-Series B*, 69, 143–161.

Yuan, M., Joseph, V. R., and Zou, H. (2009), "Structured Variable Selection and Estimation," *Annals of Applied Statistics*, 3, 1738–1757.

Zou, H. and Hastie, T. (2005), "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society-Series B*, 67, 301-320.

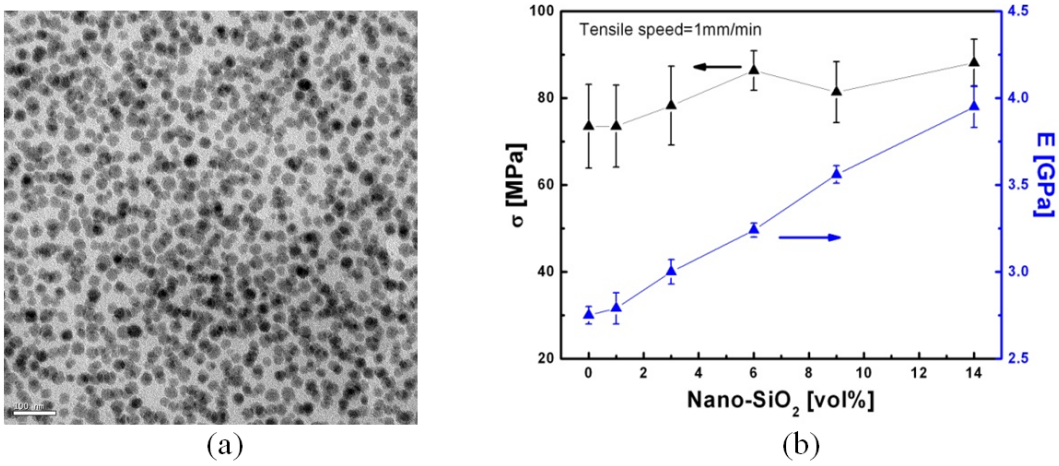
CHAPTER III

QUANTITATIVE CHARACTERIZATION AND MODELING STRATEGY OF NANOPARTICLE DISPERSION IN POLYMER COMPOSITES

3.1 Introduction

Recently, polymer-nanoparticle composite materials have attracted increasing attention due to their unique mechanical, electrical, optical, and thermal properties, see Krishnamoorti and Vaia (2002), Chapman and Mulvaney (2001), Wilson et al. (2002), and Yoon et al. (2002). These unique properties are induced by the properties of the filler particles added and by their interactions with polymer matrices (Balazs et al. (2006) and Mackay et al. (2006)). Compared to microparticles, nanoparticles have much larger surface areas that facilitate the stress transfer from the polymer matrix to the nanoparticles. In addition, the required loadings of nanoparticles in a polymer matrix are usually much lower than those of micro-fillers (Zhang et al. (2006)). As a result, when the sizes of filler particles are reduced to nanometer, dramatic improvements of material properties can be achieved (Usuki et al. (1993) and Kojima et al. (1993)). Figure 1 shows one TEM (Transmission Electron Microscope) image of polymer-nanosilica and summarizes the improved tensile modulus (the left y-axis in Figure 8(b)) and tensile strength (the right y-axis in Figure 8(b)) of polymer-nanosilica with the volume fraction of nano-SiO2 introduced (Zhang et al. (2006)).

The ultimate material properties of polymer-nanocomposites are strongly influenced by their structures, in particular, states of nanoparticle dispersion. Experimentally, significantly improved properties are usually observed from the structure possessing a uniform dispersion of nanoparticles in the polymer matrix (Zeng et al. (2008)). Taking mechanical properties as an example, a higher degree of filler dispersion has been found to provide a higher composite modulus. Besides the (1) nature of



(a) (b)

Figure 8: (a) TEM images of 14 vol.% silica/epoxy nanocomposites with an average diameter of 25 nm for nanosilica. (b) The tensile modulus and tensile strength of SiO2/epoxy nanocomposites (Zhang et al. (2006))

polymer-nanoparticle components (e.g. nanofiller, polymer, and surfactant), (2) interactions between the components and (3) volume fraction of nanofillers (as illustrated in Figure 1(b)), the polymer-nanoparticle fabrication process is the most influential determinant of nanoparticle dispersion states [Zeng et al. (2008)]. The fabrication process for polymer-nanocomposites normally starts with a sol-gel process and is followed by mechanical mixing. The sol-gel process is a chemical solution deposition technique in which the material fabrication starts either from a chemical solution (sol), or from colloidal particles to produce an integrated network (gel) (Hench and West (1990)). The sol-gel process has several advantages including the ability to regulate the proportion of organic and inorganic materials and the monodispersity of the material components (Weng et al. (2004) and Matejka et al. (1998)). However, the disadvantage posed by this process is its relatively small yield of nanocomposites with only modest mechanical properties. Larger quantities of nanocomposites with superior properties are obtained by mechanically mixing and diluting a master batch of epoxy nanocomposites produced from the sol-gel process with various nanoparticles. The mechanical mixing process significantly extends the limited varieties of nanocomposites offered by the sol-gel processing. At the same time processing conditions

during mechanical mixing also have strong influences on the states of nanoparticle dispersion and ultimately affect material properties (Mackay et al. (2006)). In spite of the important interdependence of mixing and dispersion, the underlying mechanisms by which mechanical mixing affects particle dispersion have not yet been fully understood. Particle agglomerates, which cause irregular dispersion, defects, and property degradation, often form in nanocomposites and become increasingly problematic when the filler particle content increases (Zhang et al. (2006), Maskara and Smith (2997)). Furthermore, there is no quantitative scheme to describe nanoparticle dispersion yet. Visual judgment of nanocomposite images taken by SEM (Scanning Electron Microscope), TEM or AFM (Atomic Force Microscope) is the only method in existence. Consequently, the development of uniformly dispersed nanocomposites is still largely empirical and a finer degree of control of their properties has far remained elusive (Zeng et al. (2008)). Therefore, in order to achieve nanocomposites with desired mechanical and physical characteristics, the first step should be to quantitatively model and characterize nanoparticle dispersion at the nanoscale. The modeling and characterization will be discussed in further detail throughout the course of this paper. In this paper, we propose a hierarchical modeling structure capable of integrating both existing and forthcoming physical understanding to model the dispersion of nanoparticles in the polymer-nanocomposite. Through the modeling and estimation, we will address the following key issues: (1) how to quantitatively describe the dispersion of nanoparticles in polymer-nanocomposites; (2) what is the measure for the dispersion of nanoparticles; and (3) how the proposed modeling structure can be integrated with existing and forthcoming understanding of process-structure-property relations. The objective is to provide a quantitative measure for nanoparticle dispersion and link the dispersion with process and structure characteristics. The paper is organized as follows: Section 2 discusses the modeling of nanoparticle dispersion through inhomogeneous Poisson random field together with the strategies and reasoning for choosing such a modeling; Section 3 introduces the Bayesian hierarchical modeling framework proposed for the parameter estimation; Section 4 presents case studies for both simulated data and real collected data to validate and demonstrate our proposed modeling technique; and finally Section 5 gives a summary of the proposed work and discusses the future work.

3.2 Hierarchical Modeling of Nanoparticle Dispersion at Nanoscale

Let $\boldsymbol{s}$ be the collection of sites which are regular divisions at nanoscale of the polymer-nanocomposite surface/volume and index them as $1, 2, \ldots, n$. Denote number of nanoparticles in each site s as $Y\{s\}, s \in \boldsymbol{s}$. The collection of numbers of nanoparticles at sites $\boldsymbol{s} = \{1, 2, \ldots, n\}$ is then represented as:

$$\boldsymbol{Y}(\boldsymbol{s}) = \{Y(1), Y(2), \ldots, Y(n)\}^T \tag{23}$$

What we propose is to model $\boldsymbol{Y}(\boldsymbol{s})$ as an inhomogeneous Poisson random field to develop a description of nanoparticle dispersion.

3.2.1 Strategy of Nanoparticle Dispersion Modeling

Due to the lack of physical knowledge and measurement data, Huang (2010) proposed to adopt a hierarchical modeling strategy for nanomanufacturing process modeling. The effectiveness of the hierarchical modeling has been demonstrated in nanowire growth processes (Huang (2011), Huang et al. (2010)). The proposed hierarchical strategy for nanoparticle dispersion modeling includes (1) modeling $\boldsymbol{Y}(\boldsymbol{s})$, the collection of numbers of nanoparticles at sites $\boldsymbol{s} = \{1, 2, \ldots, n\}$, as an inhomogeneous Poisson random field with site dependent intensity function $\lambda(s, \boldsymbol{\beta})$ where $\boldsymbol{\beta}$ represents the unknown parameters to be specified; (2) taking into account both process variables, nanoparticle characteristics and nanoparticle interactions for modeling the intensity function $\lambda(s, \boldsymbol{\beta})$; (3) taking the intensity function $\lambda(s, \boldsymbol{\beta})$ as a measure for nanoparticle dispersion; and (4) integrating modeling of each aforementioned components into a Bayesian framework for parameter estimation. Physical understanding will be embedded into the modeling of the intensity function and the modeling of nanoparticle characteristics and interactions. The first three parts of the strategy will be discussed in this subsection, while the Bayesian modeling framework will be discussed in Section 3.

3.2.1.1 Modeling $\boldsymbol{Y}(\boldsymbol{s})$ as an inhomogeneous Poisson random field

The concept of modeling nanoparticle dispersion through inhomogeneous Poisson random field, which is a direct extension of Poisson process from one dimension to

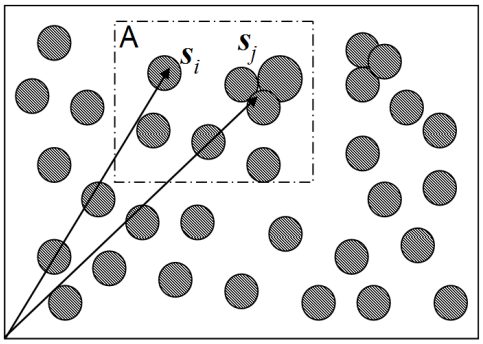


Figure 9: Inhomogeneous Poisson random field model for particle dispersion.

two or higher dimensions, stems from the observation that the number of nanoparticles in each unit area/volume and its variability at nanoscale represent the uniformity of nanoparticle distribution in polymer matrices. For simplicity, let us take a 2D model for example.

Suppose Figure 9 shows a cross-section view of a nanocomposite material at a fine scale. Clearly, the number of nanoparticles/clusters in a unit area or volume varies from region to region, and it reveals the degree of uniformity of nanoparticle dispersion on the surface observed.

Therefore, modeling nanoparticle dispersion is the same as modeling numbers of nanoparticles in sites $\boldsymbol{s} = \{1, 2, \ldots, n\}$ which are regular divisions of the nanocomposite surface/volume and have relatively small sizes. Consequently, similar to modeling event occurrences through the inhomogeneous Poisson process with a time dependent occurrence rate, modeling $\boldsymbol{Y}(\boldsymbol{s})$ as an inhomogeneous Poisson random field with site-dependent intensity function $\lambda(s, \boldsymbol{\beta})$ should also be appropriate to establish the dispersion model for nanoparticles.

3.2.1.2 Intensity function modeling of the inhomogeneous Poisson random field

Modeling of the intensity function $\lambda(s, \boldsymbol{\beta})$ includes two major components. One is the linear regression part, incorporating the effects of various process variables (e.g. mixing time, temperature, and volume fraction of added nanoparticles) and characteristics of nanoparticles/clusters (e.g. the effective size, which refers to the diameter

of a nanoparticle or the size of a cluster consisting of several nanoparticles, such as s_j shown in Figure 9). The other part is a random field $\Psi(\boldsymbol{s})$ with a special covariance structure to characterize inter-particle interactions. One potential problem with the above modeling strategy is that some of the process variables and nanoparticle characteristics are hard to quantify due to process uncertainties. For example, during the mixing process, small transient temperature variations may arise locally because of the heat generated from friction and the shear forces of the viscous fluid. A given volume element, such as region A in Figure 9, may contain several particles or clusters, making it difficult to choose an effective particle size for the intensity function. Furthermore, different process variables or nanoparticle characteristics may also have mutual dependence. For instance, the effective particle sizes are affected by temperature and pH values. The regression part will become too complex if all possible interactions among process variables and nanoparticle characteristics are directly modeled as components of it. Therefore, we propose to model process variables and nanoparticle characteristics individually if necessary and adopt a hierarchical modeling structure for their integration. Detailed development of such integration will be discussed in following sections.

3.2.1.3 Intensity function as a measure for nanoparticle dispersion

Although the intensity function is not a direct measure of the dispersion performance, it is a good index at fine scale to represent the nanoparticle dispersion state for the following reasons:

- As the length scale goes to nanoscale, the number of nanoparticles governed by the intensity function in each unit area represents the uniformity of nanoparticle dispersion on the whole nanocomposite.
- The intensity model considers effective particle/cluster sizes. The size distribution across sites indicates whether there exist agglomerates or clusters.
- Physical interaction among nanoparticles is incorporated in $\Psi(\boldsymbol{s})$. Deviation from the intended covariance structure could also indicate inappropriate dispersion.

Hence, we propose the intensity function as a quantitative measure of nanoparticle dispersion at nanoscale to facilitate control and diagnosis of the nanocomposite fabrication process in the future.

3.2.2 Modeling of Nanoparticle Dispersion

As discussed previously, we denote $\boldsymbol{Y}(\boldsymbol{s}) = \{Y(1), Y(2), \ldots, Y(n)\}^T$ to be the total numbers of nanoparticles in the areas of sites $\boldsymbol{s} = \{1, 2, \ldots, n\}$ which are regular divisions of the polymer-nanocomposite surface/volume. We model $\boldsymbol{Y}(\boldsymbol{s})$ as an inhomogeneous Poisson random field with an intensity function $\lambda(s, \boldsymbol{\beta})$. That is:

$$\Pr\{Y(s) = m\} = exp^{-\lambda(s,\boldsymbol{\beta})} \frac{\{-\lambda(s, \boldsymbol{\beta})\}^m}{m!}, \tag{24}$$

where $m = 0, 1, \ldots,$ and $n = 0, 1, \ldots, n$.

For the intensity function $\lambda(s, \boldsymbol{\beta})$, if we let $\boldsymbol{X}$ represent the collection of the influential process variables (mixing time, temperature, density of the master batch of nano-SiO2/epoxy, rotation speed, pH value, and volume fraction of added nanoparticles etc.) and nanoparticle characteristics (effective size of nanoparticles/clusters), $\boldsymbol{\beta}$ be the corresponding unknown regression coefficients and $\Psi(\boldsymbol{s})$ as the random field to capture inter-particle interactions, then $\lambda(s, \boldsymbol{\beta})$ can be represented as:

$$\lambda(s, \boldsymbol{\beta}) = \exp \boldsymbol{X}^T \boldsymbol{\beta} + \Psi(\boldsymbol{s}) \tag{25}$$

That is the log of the intensity function $\lambda(s, \boldsymbol{\beta})$ is a linear regression of $\boldsymbol{X}$ plus a correlated error term $\Psi(\boldsymbol{s})$, in contrast to the independent noises in ordinary linear regressions.

Due to process uncertainties, process variables or nanoparticle characteristics will be represented by probability distributions with model parameters ξ (e.g.$\mu(d)$ and $\sigma()$ for different size: d). And for each element in ξ, there may still be a probability distribution for it involving parameters $\boldsymbol{\theta}$. The configuration of these probability distributions will be correlated with physical process knowledge and will be defined only when necessary. Denote g, f to be corresponding possible distribution functions for a process or characteristic variable $X \in \boldsymbol{X}^T$ and parameter in ξ, then our model states:

$$X|\xi \sim g(\xi|\boldsymbol{\theta}), \text{and} \xi|\boldsymbol{\theta} \sim f(\boldsymbol{\theta}) \tag{26}$$

The correlated error terms $\Psi(\boldsymbol{s})$ will be modeled as a Gaussian Markov random field (GMRF), which defines the neighbors for each site and the conditional distribution of $\Psi(\boldsymbol{s})$ given the values of its neighboring sites, to capture the hard-to-model interaction force field among nanoparticles. The mathematical form of a GMRF is represented as follows:

$$\Psi(\boldsymbol{s})|\Psi(\boldsymbol{s}'), s \neq s' \sim N(\mu, \sigma^2(s)), \tag{27}$$

where $s, s' \in \{i, 2, \ldots, n\}$, $\mu = \sum b(s, s'')\Psi(s''), s'' \in \{\text{neighbors of} s\}$, and $\sigma^2(s)$ is the unknown conditional variance of $\Psi(\boldsymbol{s})$. Equation (27) states that the conditional distribution for each $\Psi(\boldsymbol{s})$ is a normal distribution with parameters only related to its neighboring sites.

One advantage of modeling through GMRF is that GMRF provides unified modeling representation for diverse correlation structures. This property is especially important in our case, where there is limited knowledge about the correlation structure and hence no parametric covariance functions can be derived. Besides, there are fast and efficient algorithms for transforming between GMRF and commonly used covariance functions and force fields. Hence, we can easily integrate GMRF with existing and forthcoming physical knowledge. For example, we can link it with the commonly used Lennard-Jones (L-J) potential functions [10] to characterize the interaction potential among nanoparticles. By performing the transformation between L-J potential and GMRF, not only can we make use of prior knowledge of L-J potential to get a better initial setting, but we can also gain more understanding of the process-interaction relations when statistical estimation from a real collected data is obtained. Examples of linking GMRF with L-J potential function will be demonstrated in Section 4 during the case studies.

3.3 Bayesian Hierarchical Framework for Parameter Estimation

Based on our modeling strategies presented in Section 2.1, we integrate the models specified by Equations (23)-(27) into a Bayesian hierarchical modeling framework for parameter estimation. Figure 10 depicts the framework, together with its comparison to classical hierarchical models (Rubin (1980)). We borrow the graphical representation from Huang (2011).

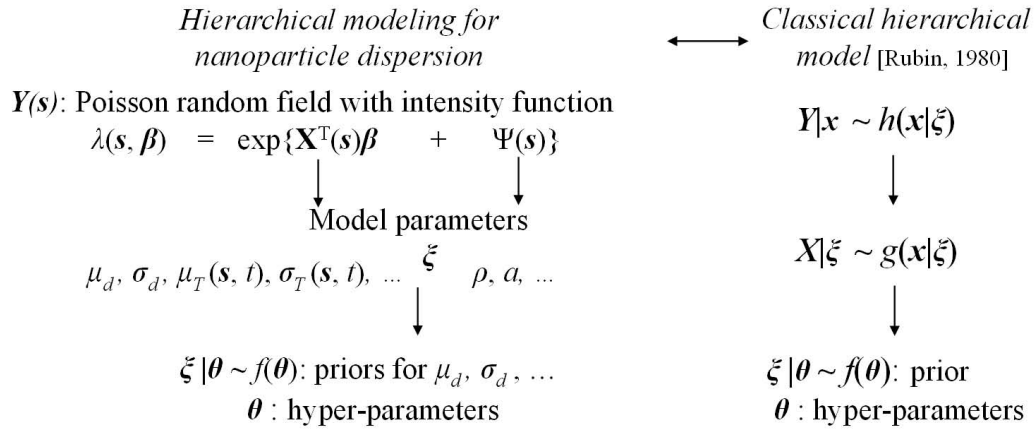


Figure 10: Bayesian hierarchical modeling framework (Huang (2011)).

Three sources of data, namely process data, product data, and prior physical knowledge for hyper-parameters $\boldsymbol{\theta}$ constitute the model inputs. The process data may include process parameters that can be quantified like pH value, and volume fraction of added nanoparticles, etc. Product data comprise the number of nanoparticles in each site counted from SEM or TEM images. Prior physical knowledge for hyper-parameters $\boldsymbol{\theta}$ will be used to set initial distributions.

There are two major advantages of adopting the Bayesian hierarchical framework. (1) Offline SEM or TEM inspection is extremely time-consuming and only provides information on a very tiny surface area of the nanocomposite. Bayesian hierarchical framework can effectively estimate large number of parameters with limited data. (2) Normally no analytical solution exists for the hierarchical model specified through equations (23)-(27). While traditional maximum likelihood estimation is hard to implement here, the Bayesian hierarchical structure allows using Markov Chain Monte Carlo (MCMC) simulation to obtain the model parameter estimation.

Given process and product data and prior distributions for hyper-parameters $\boldsymbol{\theta}$, conditional distributions of $\boldsymbol{\theta}$ under available settings can be obtained. By sequentially drawing samples via MCMC method, we could ultimately estimate $\boldsymbol{\theta}$, ξ and $\boldsymbol{\beta}$ in the model (Figure 10). Consequently, the process-dispersion relations can be estimated based on real collected data. Dispersion of nanoparticles can thus be predicted

under various process conditions. Finer degrees of control of nanocomposite properties will become possible based on the proposed quantitative modeling of nanoparticle dispersion.

3.4 Case Studies

Case studies are conducted to illustrate the proposed nanoparticle dispersion modeling approach. First, a simulation study is used to demonstrate the modeling procedure and validate the parameter estimation algorithm. Then, real experimental data will be analyzed to characterize the nanoparticle dispersion under different process conditions. The MCMC is performed using WinBUGs software in this paper. In both studies, the following procedures are followed: (1) two nanoparticle dispersion images are obtained with different parameter settings of a manufacturing process through the simulation or the real fabrication processes; (2) the images analyzed in simulation study are divided into 225 grids and the ones obtained from real fabrication are divided into 196 grids due to the image size of collected samples; (3) model parameter estimation and nanoparticle dispersion prediction are obtained through MCMC technique with WinBUGs; and (4) evaluation and discussion of obtained results are provided in both case studies.

3.4.1 Simulation Study and Discussion

The simulation dataset is designed to mimic the nanoparticle dispersion images taken from the nanocomposite specimen manufactured under different mixing processes. We assume the nanoparticle dispersion follows an inhomogeneous Poisson random field with intensity function as defined in Equations (24)-(26) of Section 3.2. We also suppose the nanoparticle interaction to follow modified L-J potential as following:

$$V(r) = 4\rho\{(\frac{a}{r+r_0})^6 - (\frac{a}{r+r_0})^{12}\}. \tag{28}$$

In Equation (28), r for the distance between any two different sites on the same image and $r_0^6 = 2(\rho + \sqrt{\rho^2 - 1})|a|^6$ so that $V(0) = 1$. Parameters used in simulation include volume fraction of nanofillers (a process variable), effective size of nanoparticles/clusters which is assumed to follow a normal distribution for both two images, and regression coefficients (β_0) as the ground intercept, (β_1) as the coefficient for

volume fraction, and (β_2) as the coefficient for effective particle/cluster size). Please see Table 1 for the summary of specified parameter values. The simulated images are drawn as Figures 11(a) and (b).

Table 1: Summary of specified $\boldsymbol{\beta}$ and ξ for dataset generation in the simulation study

Parameters	Regression Coefficients ($\boldsymbol{\beta}$)			Model Parameters (ξ)				
				Volume Fraction	Nanoparticle Effective Size		Modified L-J Potential	
	β_0	β_1	β_2	ν	μ_d	σ_d^2	ρ	a
Specified Values	2.10	1.60	0.07	(0.02, 0.15)	25 nm	4 nm^2	2.50	1.50n

As mentioned before, we follow the modeling procedures in Section 3.2, and re-estimate these model parameters through MCMC technique in WinBUGs. The non-informative priors of regression coefficients ($\boldsymbol{\beta}$) are set as: β_0, β_1, and $\beta_2 \sim$ Uniform$(0, 50)$ in this hierarchical model. Besides, we ran three Markov chains simultaneously with different initial values which are far apart from each other. The first 30000 runs are truncated to reduce the parameter estimation inaccuracy, which is so called "burn-in process", and the following 90000 iterations are used to compute posterior statistics of the interested parameters. The obtained parameter estimations are summarized in Table 2, where "95 % C.I." refers to the 95 % confidence interval (e.g. Type I error equals to 5 %).

Based on Table 2, it is clear that the estimated confidence intervals for regression coefficients ($\boldsymbol{\beta}$) cover the designated values in Table 1. Moreover, we model nanoparticle interaction as first order CAR model and re-estimate ρ and a in the modified L-J potential. The estimation results of regression coefficients and L-J potential parameters support the effectiveness of the proposed estimation algorithm. Table 2 also shows that if the volume fraction of the nanoparticles is higher, the estimated mean value of the intensity function is higher, which matches our intuition.

The four panels of Figure 11 are used to compare the number of dispersed nanoparticles from generated simulation dataset with the ones obtained by the model predictions. In each panel, the image plot shows the number of dispersed nanoparticles associated with each grid, s, and the x- and y-axes are the indices for different grids. To be more specific, Figures 11(a) and 11(b) are numbers of nanoparticles counted

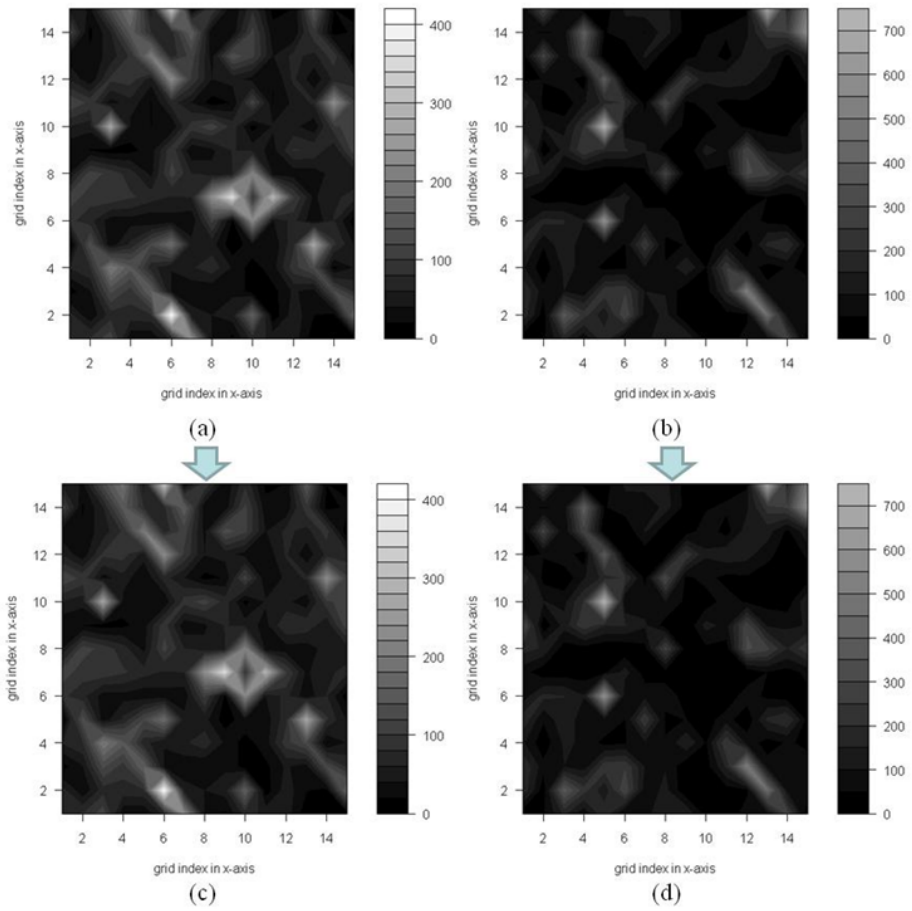


Figure 11: Comparison of numbers of nanoparticles dispersion between simulated and predicted samples: (a) simulated image with 2 vol.%, (b) simulated image with 15 vol.%, (c) predicted image with 2 vol.%, and (d) predicted image with 15 vol.% nanoparticels introduced.

Table 2: Bayesian estimates via MCMC for simulation data study

Parameters	Mean	Std. Dev.	Median	95% C.I.
β_0	1.097	0.954	0.665	[0.025, 3.231]
β_1	1.772	0.197	1.774	[1.374, 2.160]
β_2	0.1116	0.038	0.129	[0.026, 0.155]
ρ	2.180			
a	1.184			

from simulated specimens; while the predicted numbers of nanoparticles for these two specimens are plotted as Figures 11(c) and 11(d), which are actually equal to the estimated values of intensity function. The left column in Figure 11 stands for nanocomposite with 2 vol.% nanoparticles while the right column is nanocomposite with 15 vol.% nanoparticles. Overall to say, the patterns of nanoparticle dispersion look quite similar between the simulated data and the prediction results.

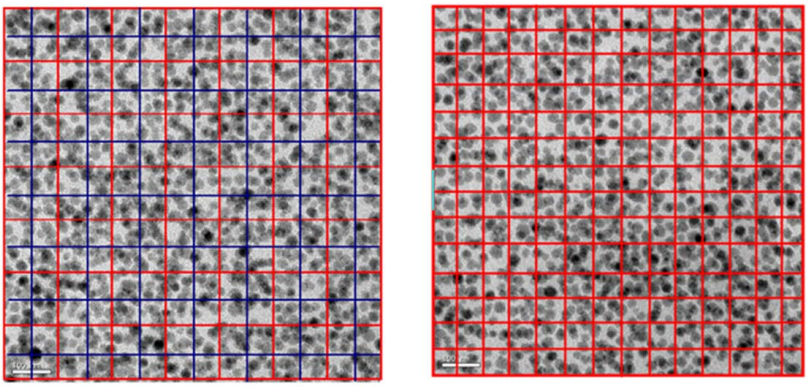
3.4.2 Real Image Data Study and Discussion

Figures 12(a) and 12(b) shows two images taken from real nanocomposite surfaces after mixing process and the size of each photo is about 700×700 nm^2. Each of them is further divided into 196 sites in total which is determined by the appropriate cross-section size suggested by the domain experts. We use these two images to develop the quantitative model for nanoparticle dispersion characterization.

Based on domain experts' knowledge, the influential process variables are similar to the ones we used in the simulation case study including volume fraction, ν, and effective particle/cluster size, d. Since the temperature is maintained at the same degree throughout the whole mixing process, it is eliminated from the analysis.

The hierarchical modeling structure and the parameter estimation process are carried out similarly as what have been done in the simulation study. The non-informative priors of regression coefficients ($\boldsymbol{\beta}$) are set as: $\beta_0 \sim \text{Uniform}(0, 10), \beta_1 \sim \text{Uniform}(0, 1e^5)$, and $\beta_2 \sim \text{Normal}(0, 1e^5)$ in the hierarchical model. Similarly, we adopt the first order CAR model in this analysis to model the nanoparticle interactions.

Table 3 summarizes the parameter estimation results obtained from the three



(a) (b)

Figure 12: Two TEM images with grid line taken from the silica/epoxy nanocomposites with different amount of nanoparticles introduced: (a) 9 vol.% (b) 14 vol.%.

different Markov chains gone through burn-in process. Here the regression coefficients are consistent with the notations defined earlier in the simulation study. Based on the estimated regression coefficients in Table 3, it is clear that increasing volume fraction of nanofillers will increase the density of dispersed nanoparticles. Also, the site with a larger effective nanoparticle/cluster size will tend to have less number of nanoparticles dispersed.

Table 3: Bayesian estimates via MCMC for real experimental images (silica/epoxy nanocomposite with 9 vol. % and 14 vol. % induced nanoparticles)

Parameters	Mean	Std. Dev.	Median	95% C.I.
β_0	0.929	0.577	0.862	[0.064, 2.173]
β_1	4.399	1.156	4.395	[2.135, 6.679]
β_2	-0.011	0.023	-0.009	[-0.060, 0.025]
ρ	2.800			
a	2.285			

To clearly represent the counting quantity $\boldsymbol{Y}(\boldsymbol{s})$ at a fine scale, Figures 13(a) and 13(b) plot the counted number of nanoparticles dispersed in each site on the two nanocomposite images, Figures 12(a) and 12(b); and Figures 13(c) and 13(d) are the predicted number of dispersed nanoparticle through WinBUGs. Note that images on

the left in Figure 13 are the images for 9 vol.% silica/epoxy nanocomposites while the images on the right are those for 14 vol.% silica/epoxy ones. The predictions are the estimated values of intensity functions.

One thing worthy to mention is that since we adopt the inhomogeneous Poisson random field as the key component of characterization model, the prediction actually represents the "expected" nanoparticle dispersion states in a general probabilistic realization point of view. Moreover, one interesting observation is that the 95% confidence intervals of intensity function cover the truly counted numbers of nanoparticles in most of the cases, as shown in Figure 14. The solid black dots stand for the counted numbers from the real image data, while the empty triangles are the predictions whose 95% confidence bounds are depicted by gray lines in Figure 14.

Other than looking at Figure 14 to learn the model prediction performance, we also calculate (1) mean of estimated intensity functions among all sites in the same image, (2) mean of standard deviations of intensity functions among all sites in the same image, (3) mean of 95% confidence intervals of intensity function among all sites in the same image, and (4) mean of the counted numbers of nanoparticles among all sites in the same image, see Table 4. It is clear that the mean of true observational numbers of dispersed nanoparticles counted from each TEM image is within the mean of 95% confidence intervals of λ. This gives another example to support the effectiveness of our proposed model structure and parameter estimation strategy.

Table 4: Predictive ability of the proposed model

Vol. % of nanofillers for different images	Mean of λ prediction	Mean of Std. Dev. of λ prediction	Mean of 95% C.I. of λ	Mean of truly counted numbers
9 %	2.891	0.448	[2.107, 3.861]	2.883
14 %	3.597	0.540	[2.648, 4.764]	3.602

3.5 Summary

Nanoparticle dispersion plays a crucial role in determining the mechanical properties of polymer nanocomposites. In this work we develop a quantitative measure to effectively describe nanoparticle dispersion based on microscope measurement data.

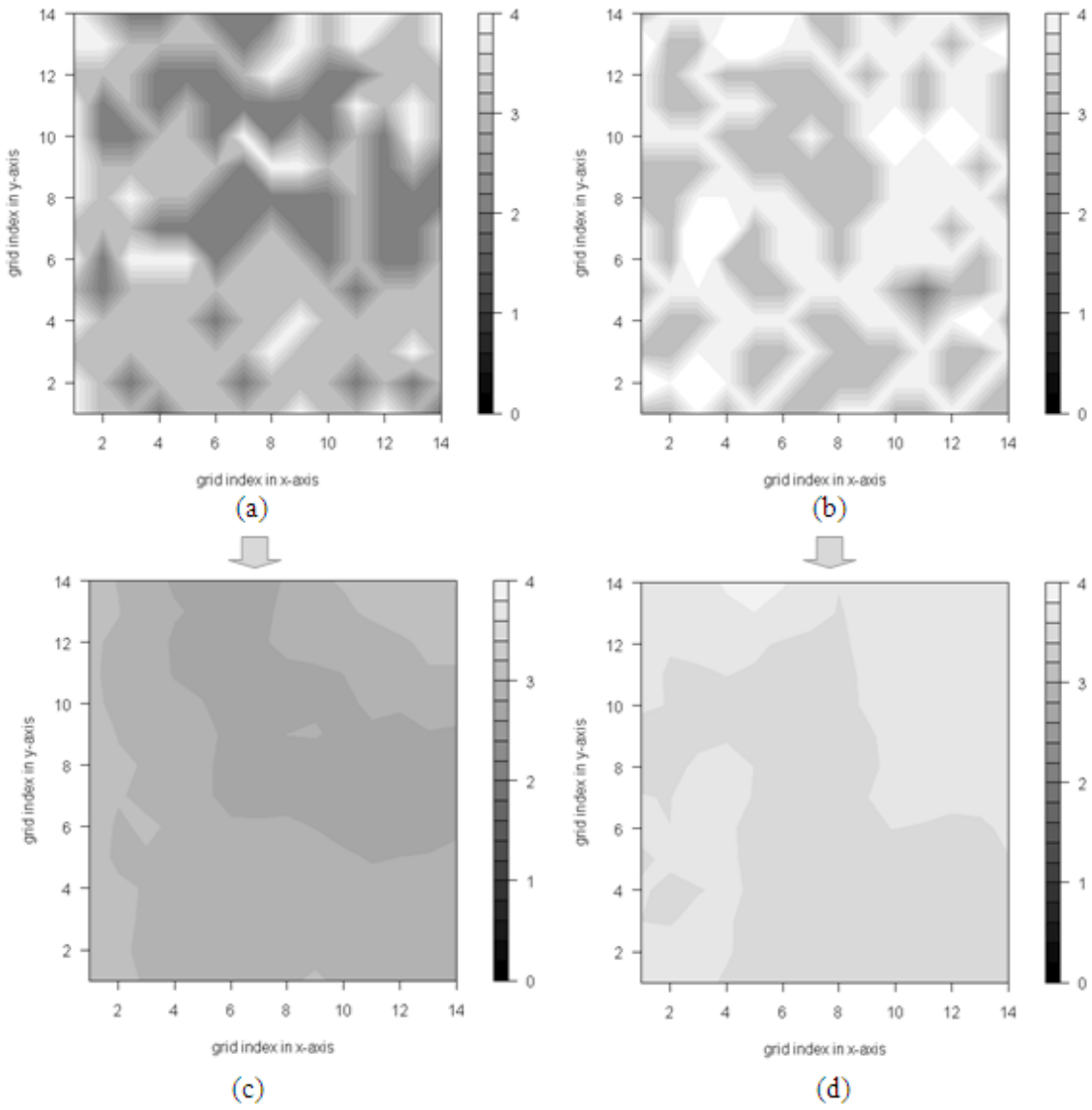
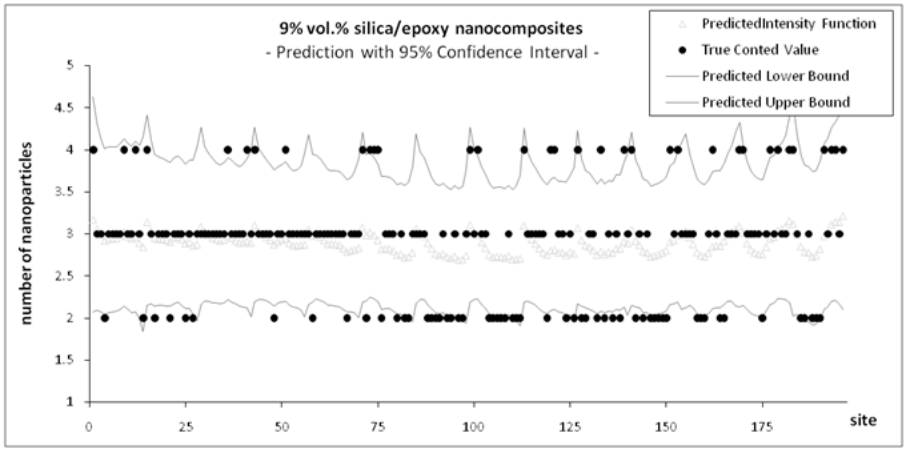
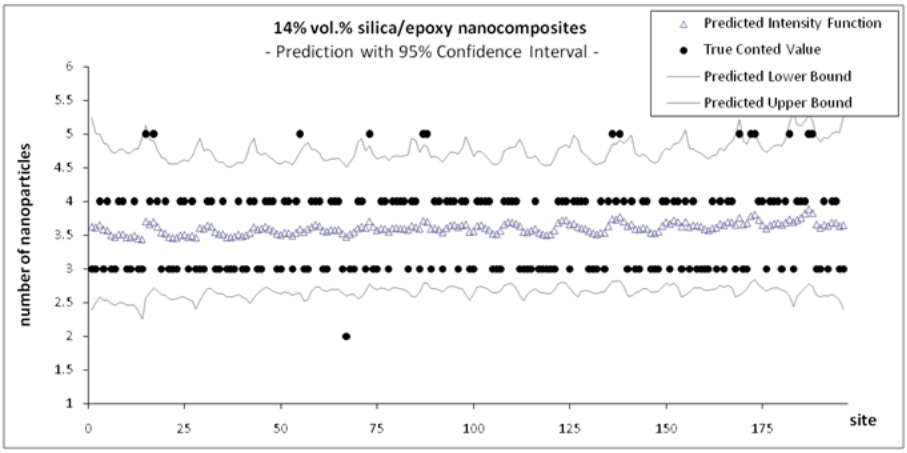


Figure 13: Comparison of counted nanoparticle dispersion states between true image and model prediction: (a) true image with 9 vol.%, (b) true image with 14 vol.%, (c) prediction with 9 vol.%, and (d) prediction with 14 vol.% silica/epoxy nanocomposites.



(a)



(b)

Figure 14: Comparison of observation and prediction of nanoparticle dispersion states: (a) 9 vol.% silica/epoxy nanocomposites, (b) 14 vol.% silica/epoxy nanocomposites.

The nanoparticle dispersion is characterized by the intensity function of an inhomogeneous Poisson random field. The intensity function is a mixture of the linear regression model and the Gaussian Markov random field. To be more specific, the linear regression model is constructed with process variables and nanoparticle characteristics while the Gaussian Markov random field is approximated from engineering knowledge of particle interaction force. Unlike the conventional modeling methods that usually rely only on pure physical laws or statistical data-driven techniques; the proposed model integrates both nanomanufacturing domain knowledge and statistical data analysis to provide a better characterization of nanoparticle dispersion states. Both simulation and experimental data analysis show the effectiveness of the proposed method.

Since the proposed model link process variables with final nanocomposite structure properties, it provides a basis for process monitoring, root cause diagnosis and active process control, thus deserves further investigation in the future.

3.6 References

Balazs, A.C., Emrick, T., and Russell, T.P. (2006), "Nanoparticle Polymer Composites: Where Two Small Worlds Meet," *Science*, 314, 1107-1110.

Chapman, R., Mulvaney, P. (2001), "Electro-optical shifts in silver nanoparticle films," *Chemical Physics Letters*, 349, 358 -362.

Hench, L.L., West, J.K. (1990), "The Sol-Gel Process," *Chem. Rev.*, 90, 33-72.

Huang, Q., Wang, L., Dasgupta, T., Zhu, L., Sekhar, P.K., and, Bhansali, S., An, Y., 2010, "Statistical Weight Kinetics Modeling for Silica Nanowires Growth Catalyzed by Pd Thin Film," *IEEE Transactions on Automation Science and Engineering*, DOI: 10.1109/TASE.2010.2070493

Huang, Q. (2011), "Physics-Driven Bayesian Hierarchical Modeling of Nanowire Growth Process At Each Scale," *IIE Transactions*, Vol. 43, pp. 1-11.

Kojima Y., Usuki, A., Kawasumi, M., Okada, A., Fukushima, Y., Kurauchi, T., and Kamigaito, O. (1993), "Mechanical properties of nylon 6-clay hybrid," *Journal of Materials Research*, 8, 1185-1189.

Krishnamoorti, R. and Vaia, R.A. (2002), Polymer nanocomposites, 804. Washington, DC: ACS.

Mackay, M. E., Tuteja, A., Duxbury, P. M., Hawker, C. J., Horn, B. V., Guan, Z., Chen, G., and Krishnan, R. S. (2006), "General Strategies for Nanoparticle Dispersion," *Science*, 311, 1740-1743.

Maskara, A., Smith, D.M. (1997), "Agglomeration during the drying of fine silica powders .2. The role of particle solubility," *Journal of the American Ceramic Society*, 80(7), 1715-1722.

Matejka, L., Dusek, K., Noga, J. (1998), "Formation, structure and mechanical properties of organic-silica hybrid networks," *Wiley Polym Networks Group Rev Ser*, 1, 301-311.

Rubin, D.B. (1980), "Using Empirical Bayes Techniques in the Law School Validity Studies," *Journal of the American Statistical Association*, 75, 801-816.

Usuki, A., Kojima, Y., Kawasumi, M., Okada, A., Fukushima, Y., Kurauchi, T., and Kamigaito, O. (1993), "Synthesis of nylon 6-clay hybrid," *Journal of Materials Research*, 8, 1179-1184.

Weng, W.H., Chen, H., Tsai, S.P., Wu, J.C. (2004), "Thermal property of epoxy/SiO2 hybrid material synthesized by the sol-gel process," *Journal of Applied Polymer Science*, 91(1), 532-537.

Wilson, O., Wilson, G.J., Mulvaney, P. (2002), "Laser writing in polarized silver nanorod films," *Advanced Materials*, 14, 1000-1004.

Yoon, PJ., Fornes, T.D., Paul, DR. (2002), "Thermal expansion behavior of nylon 6 nanocomposites," *Polymer*, 43, 6727-6741.

Zhang, H., Zhang, Z., Friedrich, K., Eger, C. (2006), "Property improvements of in situ epoxy nanocomposites with reduced interparticle distance at high nanosilica content," *Acta Materialia*, 54, 1833-1842.

Zeng, Q.H., Yu, A.B., Lu, G.Q. (2008), "Multiscale modeling and simulation of polymer nanocomposites," *Progress in Polymer Science*, 33, 191-269.

CHAPTER IV

ENHANCING THE FORCE MODEL OF THE LASER ASSISTED MICRO MILLING (LAMM) PROCESS

4.1 Introduction

There is an increasing demand to machining micro-scale features in materials such as hardened mold and die steels (60-65 HRc), see Aramcharoen et al. (2008), Bissacco et al. (2005), Bissacco (2006) and Melkote et al. (2009). To overcome the shortcomings of low material removal rates, rapid tool wear/failure, and poor part feature accuracy when machining hard materials with microscale features, hybrid processes like LAMM have been developed. This process seeks to reduce the strength of the hard workpiece material through localized thermal softening with the aid of laser irradiation.

Singh and Melkote (2005, 2007) demonstrated the use of laser heating to induce localized thermal softening in a micro grooving process. Melkote et al. (2009) successfully designed and built a LAMM setup capable of producing three dimensional features in hard materials. In general, laser assist can increase material removal rate along with reduce tool wear and cutting force needed. However, it will cause burr formation and poor surface finish due to the thermal softening effect on the workpiece. This leads to an important process design problem of how to determine the combination of process variables in LAMM which leads to higher material removal rate by reducing the cutting forces, while keeping the burr formation minimum and improving surface finish quality. To fulfill this need, an accurate force model is required to understand the effects on cutting forces caused by different machining and laser parameters as well as the estimation of temperature distribution on the thermally affected zone.

Very little work has yet been done on cutting force model development in laser assisted micro machining process. Singh and Melkote (2007) developed a force model to predict cutting forces in laser assisted micro grooving process and used the model to optimize the process parameters, see Singh et al. (2011). Kumar (2011) first

develop the physics based model to predict cutting forces in LAMM process; in this work, LAMM experiments are carried out on 52100 bearing steel (62 HRc), over a range of feed rates and laser powers to calibrate and validate the force model.

4.2 Force Model

Kumar (2011) proposed the cutting force model by integrating the four key elements together: a thermal model for laser heating (Carslaw and Jaeger (1959)), a thermal model for the prediction of temperature rise due to the cutting process (Loewen and Shaw (1954)), a shear angle model for chip formation (Wright(1982)), a material model for flow strength (Guo and Liu (2002)), a mechanistic model for the milling process (Budak et al. (1996)) and a runout model (Marsh(2008)). Among them, runout model plays an extremely important role to capture the discrepancy between physical force measurements and predictions from original physics based model. By considering the runout phenomenon happened in the experiments, the force model does a better job in predicting cutting force in LAMM. One challenge in such force model development is how to estimate the unknown parameters in LAMM force model, the friction angle β and the runout parameters $(A, B, \text{and} c)$ simultaneously to improve model predictive ability. The details of the proposed parameter estimation procedure will be provided in Section 4.3.

4.2.1 Overview of Cutting Force Model

The inputs of LAMM force model include laser parameters (power, spot size and the laser-cutting tool distance) and the machining parameters (diameter of the cutting tool, feed, spindle speed, depth of cut, and runout parameters). The radial and tangential forces produced in the laser assisted micro milling process are the outputs of force model. Note that the radial and tangential forces are appropriately transformed to the desired dynamometer coordinate system to validate the predicted forces against the measured cutting forces. The framework of the proposed force model in Kumar (2011) is drawn in Figure 15. Among the four key modules of prediction model, we refit the material model and keep the other three modules as the same structure.

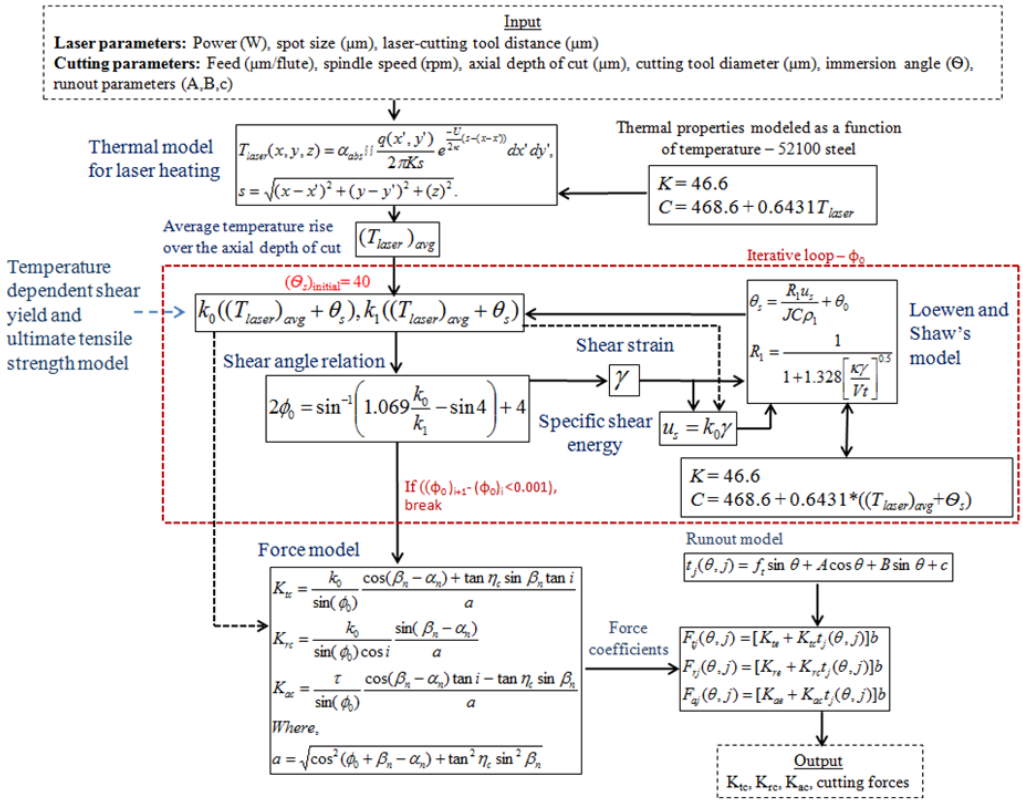


Figure 15: Flowchart of the force prediction methodology (Kumar (2011)).

4.2.2 Material Models

Unlike the material models proposed in Kumar (2011), We modify the models of shear yield and ultimate tensile strength by proposing the nonlinear models matching training data better. The variation of yield and tensile strengths with temperature is essential to calculate the shear angle in Figure 15. Guo and Liu (2002) have conducted cutting experiments and accounted for the effect of strain rate through the velocity modified temperature approach. The tensile and yield strength data for 52100 steel with respect to temperature are plotted in Figure 16. In addition, Guo and Liu (2002) have shown that 52100 steel is insensitive to strain rates typically encountered in the machining process ($\sim 10^4 - 10^5$/s). The effect of strain hardening varies with each material and appropriate models have to be incorporated to account for this effect. Two nonlinear regression models are fitted to the tensile and yield strength data points and the resulting fits are also shown in Figure 16. The R^2 values for both two models are 0.997, indicating a very good fit. The fitted models for tensile and yield strength predictions are:

$$\begin{aligned} Y_{\text{Tensile}}(T) &= 1251.98(1 + 0.00494T)e^{-(\frac{T}{328.647})^2} \\ Y_{\text{Yield}}(T) &= 1799.88(1 + 0.00494T)e^{-(\frac{T}{328.647})^2} \end{aligned} \tag{29}$$

4.2.3 Runout Model

When assuming the instantaneous uncut chip thickness with no runout, denoted as: $t_j(\theta, j) \cong f_t \sin\theta_j(z)$, the predicting tangential and radial force could be obtained by Equation (30)

$$\begin{aligned} F_{tj}(\theta, j) &= [K_{te} + K_{tc}t_j(\theta, j)]b, \\ F_{rj}(\theta, j) &= [K_{re} + K_{rc}t_j(\theta, j)]b, \end{aligned} \tag{30}$$

where b is the depth of the cut. To compare the predicted forces with the measured forces, the radial and tangential forces are transformed into the dynamometer coordinate system using the following equations:

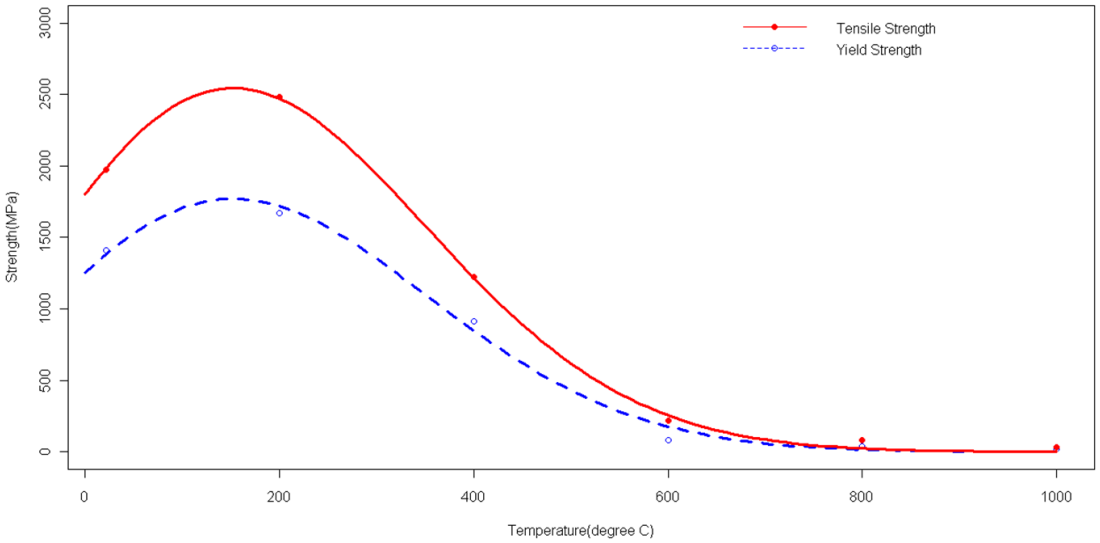


Figure 16: Temperature dependence of the yield and ultimate tensile strengths of 52100 steel (62 HR_c).

$$F_X = -F_t \sin(\theta) + F_r \cos(\theta)$$
$$F_Y = F_t \cos(\theta) + F_r \sin(\theta). \tag{31}$$

Based on Equations (30) and (31), a profile of cutting force prediction is drawn in Figure 17 along with the experimentally measured data. It can be seen that the force model does not capture the troughs and the peaks in the measured force signal well. This discrepancy is due to the effect of spindle runout combined with the cutting tool axis offset runout on the uncut chip thickness in the micro milling process. This effect can be taken into consideration using the three parameter runout model (Guo and Liu (2002)),

$$t_j(\theta, j) = f_t \sin\theta_j + A\cos\theta_j + B\sin\theta_j + c, \tag{32}$$

where f_t is the feed per tooth, θ is the immersion angle of the flute, A and B are the spindle runout parameters assuming a two lobe runout model, and c is the actual axis offset runout of the cutting tool due to center offset and tool geometric errors, see Kumar (2011). Note that the change in runout parameters of the spindle due to the cutting process is not considered here. Moreover, only a two lobe runout model is taken into account due to the difficulty in measuring the runout parameters for a

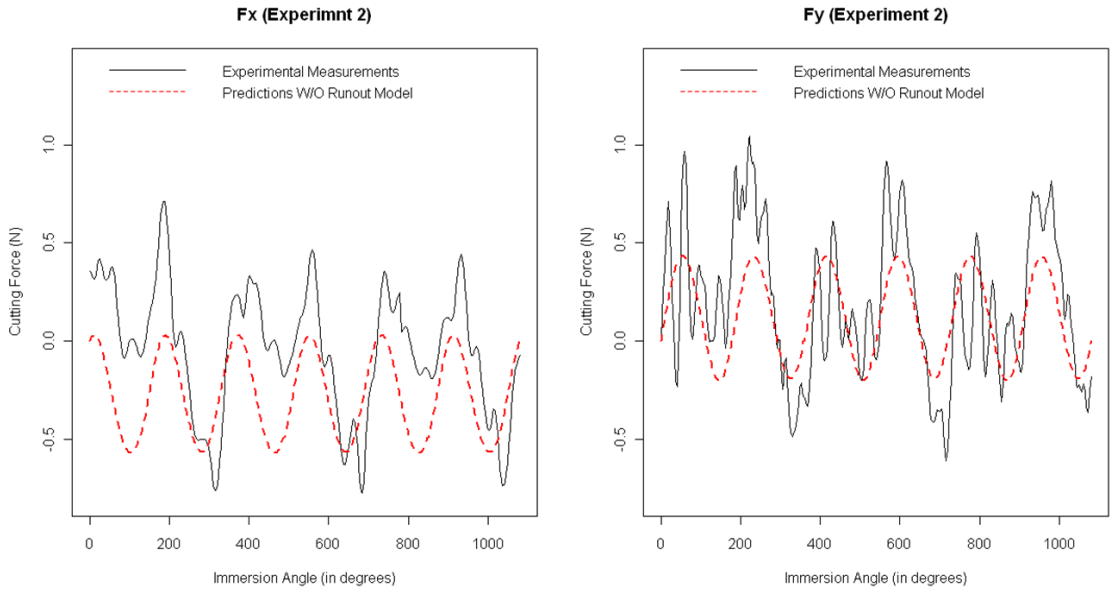


Figure 17: Model prediction of F_X over three cutter rotation cycles along with the experimental data without the runout model (feed: 6.6 um /flute, depth: 20 um, spindle speed: 50,000 rpm, laser power: 19 W, scan speed: 660 mm/min, tool diameter: 180 um, no. of flutes: 2, distance between the center of the laser beam and the edge of the cutting tool: 200 um).

higher lobe model for the micro end mills. More details regarding to runout model, please refer to Kumar (2011).

The two lobe run-out parameters A and B could be measured experimentally using two laser interferometers as shown schematically in Figure 18. Direct measurement of runout at the tool tip is very challenging at the microscale since the spot size of the laser interferometer (0.8 mm) is much greater than the diameter of the tool. In addition to directly measuring runout parameters A and B, they can be estimated indirectly from the force measurements.

The third parameter c is the actual shift of the tool axis with respect to the spindle axis as indicated in Figure 17. This error might be due to a number of factors such as the difference in the radii of the flutes, center offset and so on and hence this parameter needs to be estimated statistically. The details of the parameter estimation procedure will be given in Section 4. Note that the runout model in Equation (32) does not explicitly account for any dynamic effects. The results presented in subsequent sections indicate that the model predicts the force profile with good accuracy even with the two lobe assumption.

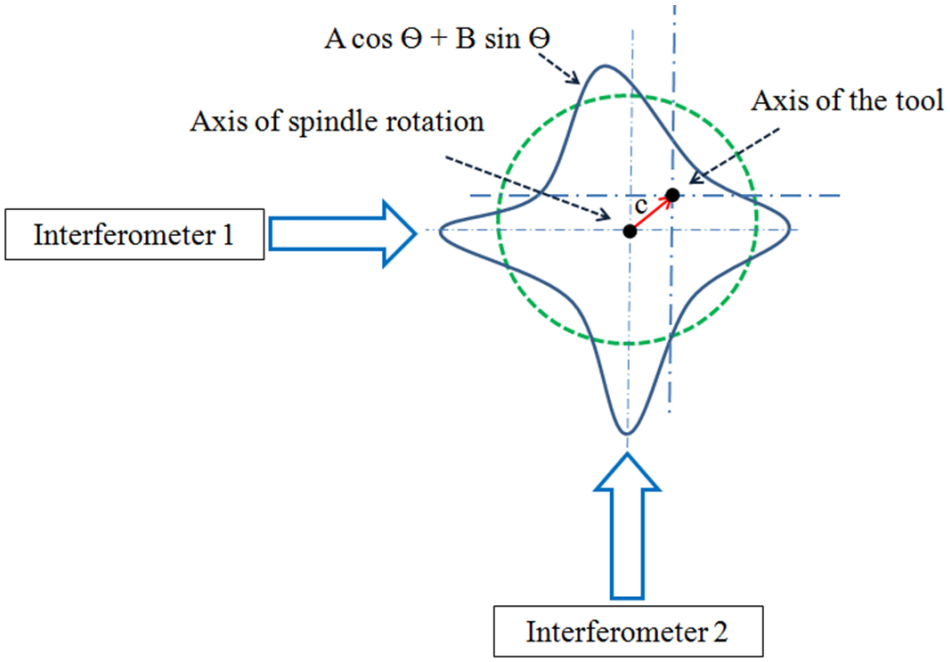


Figure 18: Schematic of the runout parameters and its measurement (Kumar (2011)).

4.2.4 Micro Milling Experiments

Bearing steel (62 HR_c) with the nominal composition was adopted as workpiece in this study. Two-flute 180?m diameter, TiAlN coated tungsten carbide square end mills (SECO JM905) were used in the experiments. A series of slot milling experiments, summarized in Table 5, were performed to calibrate the force model. See Kumar (2011) for details of experiment implementation.

Table 5: Experimental conditions for a full factorial design (Kumar (2011))

Experiment	Tool Number	Feed (um/flute)	Laser Power (W)	Axial depth of cut (um)
1	1	2.2	18	16
2	1	6.6	18	16
3	2	2.2	12	16
4	2	6.6	12	16
5	2	2.2	24	16
6	2	6.6	24	16

4.3 Calibration Methodology and Results

As discussed in the aforementioned section, there are four unknown parameters in the proposed force model: A, B, c, and β and these parameters need to be estimated based on experimental force measurements. Although A and B are possibly obtained through direct measurements in the experiment, the measurement error may influence the accuracy of A and B, and potentially lead to errors in the estimates of c and β. Hence, these parameters are estimated statistically from the force measurements to compare the results obtained from actual experimental measurements.

4.3.1 Parameter Estimation Procedure

Let $Z_X(\theta)$ and $Z_Y(\theta)$ be the measured values of forces in the X and Y directions. The measured values are affected by measurement errors in addition to the random variability affecting the process. Therefore, we can write down a statistical model as:

$$\begin{aligned} Z_X(\theta) &= F_{X_{j=1}}(\theta - a) + F_{X_{j=2}}(\theta - a) + \epsilon_X, \\ Z_Y(\theta) &= F_{Y_{j=1}}(\theta - a) + F_{Y_{j=2}}(\theta - a) + \epsilon_Y, \end{aligned} \tag{33}$$

where $\epsilon_X \sim \mathrm{N}(0, \sigma_X^2)$, and $\epsilon_Y \sim \mathrm{N}(0, \sigma_Y^2)$. Because the force measurements are collected continuously, it is difficult to locate the start of the cut during the micro milling process. Therefore, an additional parameter a is introduced to the model to estimate the angular position of engagement of the flute with the workpiece.

The parameters A, B, c, β, and a can be obtained by minimizing the sum of squares of errors: $Q(A, B, c, \beta, a) = \sum_i \epsilon_{X_i}^2 + \sum_i \epsilon_{Y_i}^2$. Substituting Equations (30), (33) and cutting coefficients formula in Figure 15 in (33). Hence, the least squares solution can be obtained explicitly in terms of β and a.

4.3.2 Results of Parameter Estimates

The five unknown parameters are estimated and the results are summarized in Table 6. Note that the β estimates of experiments 5 and 6 are different from the remaining cases. This might be due to the change in the contact conditions at higher laser powers which results in increased adhesion/rubbing of the work material with the tool, resulting in an increase in β. Hence, in the subsequent analysis, β value is assumed to be equal to the average of the estimates obtained from the first four

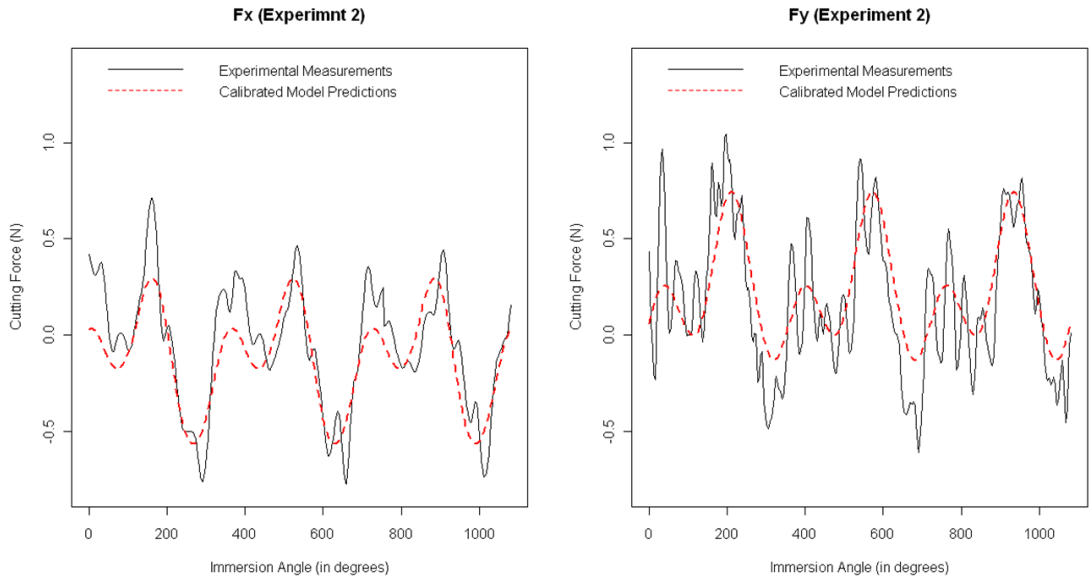


Figure 19: Comparison between the predicted and the experimentally measured cutting forces.

experiments. The measured values of the runout parameters, $A = 1.8$ um and $B = -0.5$ um compares well with the average value of the estimated values from the table ($A_{avg} = 2.9$um, $B_{avg} = 1.2$um).

Table 6: Force model parameter estimation results.

Experiment	A (um)	B (um)	c (um)	β (angle)	a (in degrees)
1	1.192	-0.254	4.078	29.11	47.368
2	3.226	-1.906	5.198	28.28	34.737
3	2.646	-0.194	1.476	28.89	34.737
4	7.009	0.881	0.982	29.11	37.895
5	2.004	-1.953	1.007	50.40	37.895
6	2.774	-3.320	1.092	48.55	31.579

The predicted force profile for experiment 2 is shown in Figure 19 where the calibrated force model captures the peak forces with reasonable accuracy. In addition, the measured and the predicted force profiles align better in Figure 19 when compared with Figure 17. This shows the usefulness of including an additional parameter a in determining the starting immersion angle of cut.

4.3.3 Uncertainty in the Predictions

The actual shift of the tool axis with respect to the spindle axis, c changes between experiments. This necessitates the need to account for the variability in the predicted cutting forces between different experiments. The value of cutting tool axis offset in each run is assumed to follow $\mathrm{N}(\mu_c, \sigma_c^2)$ where μ_c and σ_c^2 could be computed from Table 6. The variation in the predicted cutting forces is quantified by considering the variability of c in addition to the variability in the actual force measurements:

$$\begin{aligned}\mathrm{Var}(F_X(\theta)) &= \sigma_c^2(K_{tc}(\beta)\sin(\theta) - K_{rc}(\beta)\cos(\theta))^2 + \sigma_X^2,\\ \mathrm{Var}(F_Y(\theta)) &= \sigma_c^2(-K_{tc}(\beta)\cos(\theta) - K_{rc}(\beta)\sin(\theta))^2 + \sigma_X^2,\end{aligned} \tag{34}$$

Based on the experiments listed in Table 5, σ_X^2 and σ_Y^2 are estimated to be 0.069 and 0.078, respectively, which shows that the force measurement in the Y direction have larger variability when compared to the X direction. Given Equation (34), the 95% prediction interval of the cutting forces can be determined as follows:

$$\begin{aligned}\hat{F}_X(\theta) \pm 1.96\{\widehat{Var}[F_X(\theta)]\}^{1/2} &= \hat{F}_X(\theta) \pm 1.96\{\hat{\sigma}_c^2[K_{tc}(\beta)\sin(\theta) - K_{rc}(\beta)\cos(\theta)] + \hat{\sigma}_X^2\}^{1/2},\\ \hat{F}_Y(\theta) \pm 1.96\{\widehat{Var}[F_Y(\theta)]\}^{1/2} &= \hat{F}_Y(\theta) \pm 1.96\{\hat{\sigma}_c^2[-K_{tc}(\beta)\cos(\theta) - K_{rc}(\beta)\sin(\theta)] + \hat{\sigma}_Y^2\}^{1/2}.\end{aligned} \tag{35}$$

4.3.4 Validation Experiment

The following experiment is used to validate the force model: feed: 4.4 (um/flute), laser power: 18 (W) and axial depth of cut: 16 (um). Figure 20 shows the predicted cutting force profile with its 95% prediction interval along with the measured force profile. The model predictions capture the peak and trough force with acceptable accuracy and the entire profile is captured within the 95% prediction interval.

4.3.5 Model Prediction

Figures 21 and 22 show the variation of the temperature rise at the edge of the machined groove and the maximum cutting forces using different laser powers and laser-cutting tool distances, respectively. The estimation of temperature rise at all locations along the edge of the groove helps select cutting and laser scanning conditions that can reduce the thermally affected region outside the machined groove.

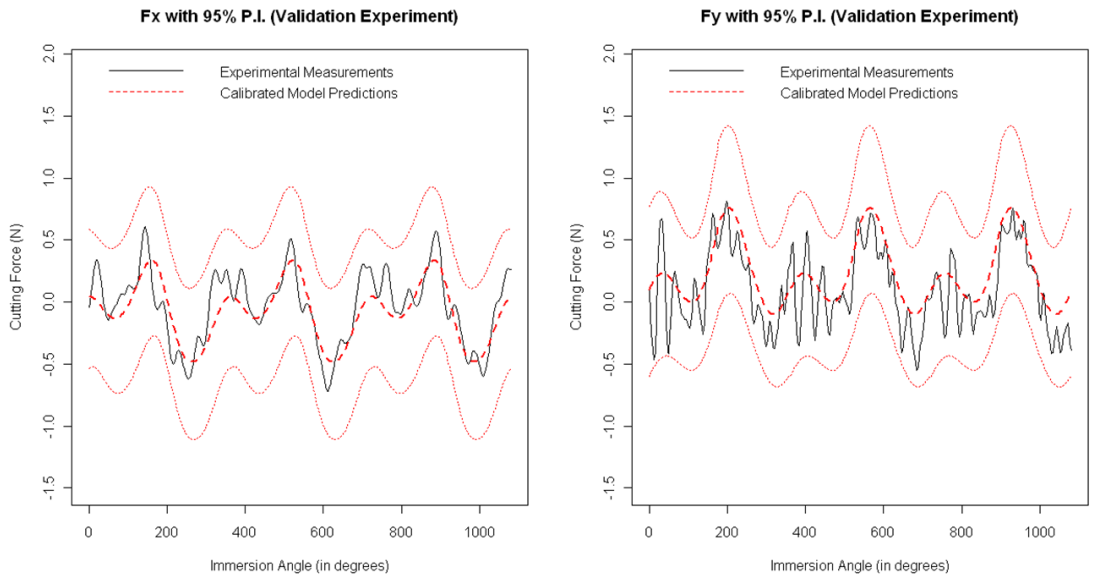


Figure 20: Predicted cutting forces with its 95% prediction interval of validation experiment.

As the laser power increases, the temperature increases resulting in a drop in the cutting forces as seen in Figure 21. Note that when the laser power is low ($5-10$ W), the cutting force is higher when compared to the without laser assist case as seen in Figure 22. This is due to the increase in yield strength with temperature as shown in Figure 16. These figures enable the users to select optimal process parameters. As an example, for a cutting tool-laser distance of 150 um, the laser power should not exceed 14 W if the temperature constraint is set to 350 C at the edge of cut. The corresponding cutting forces can be determined from Figure 22 to be 0.94 N.

4.4 Summary

A physics-based force model for LAMM that captures the effect of thermal softening due to laser heating is required to better understand LAMM process for experimental planning. The developed force model considers runout effect which brings the original force model closer to LAMM in practice. However, to incorporate runout phenomenon in force model successfully, one challenge is how to calibrate the runout parameters and model parameters simultaneously to achieve accurate model predictive ability.

In this chapter, the unknown parameters in the model such as the friction angle (β) and the runout parameters ($A, B,$ and c) are calibrated statistically with force

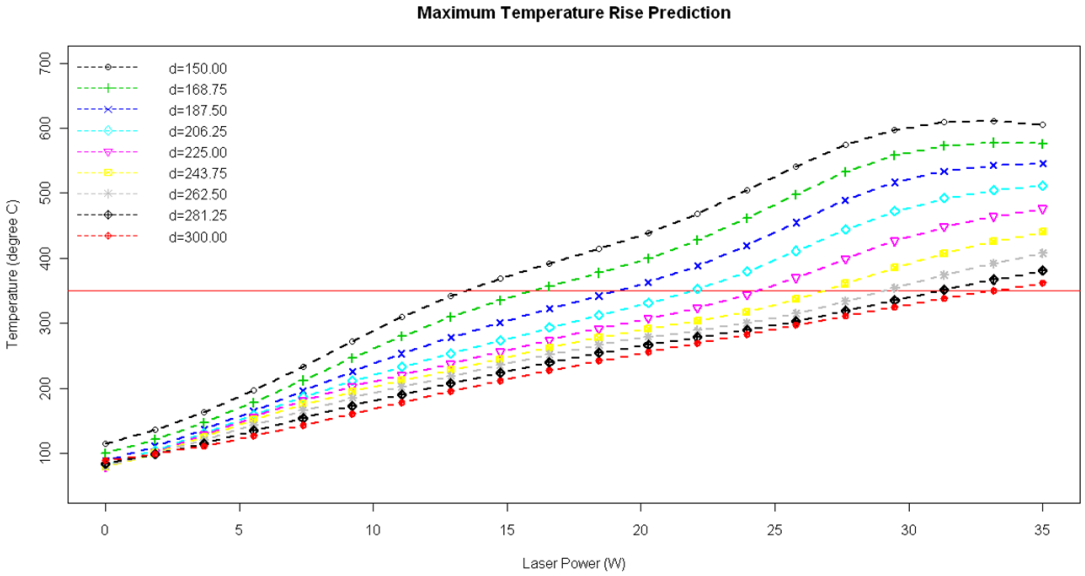


Figure 21: Variation of the predicted temperature rise at the edge of the machined groove under varying laser powers and laser-cutting tool distances.

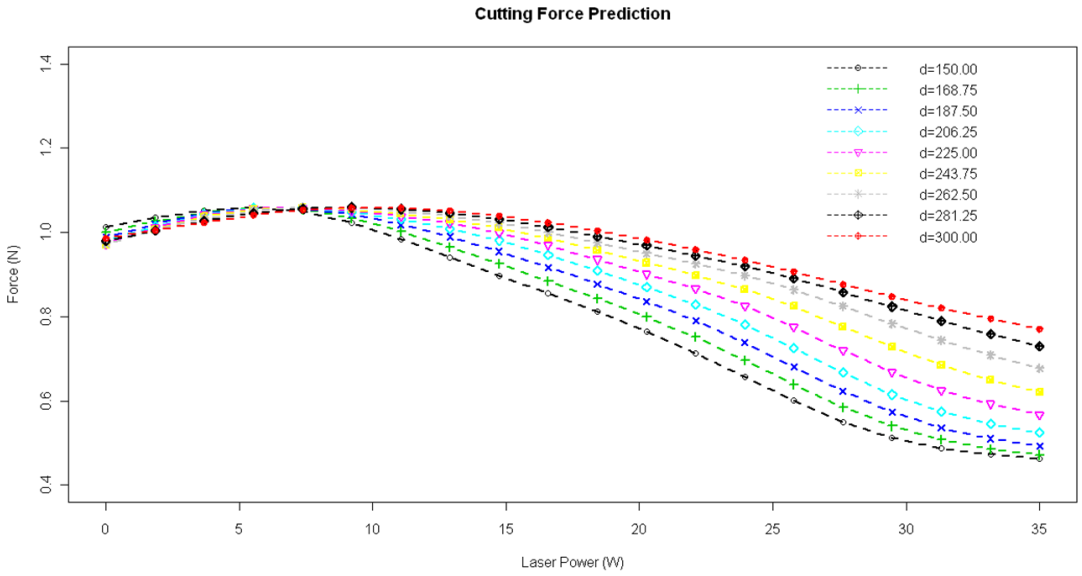


Figure 22: Variation of the cutting forces in LAMM under varying laser powers and laser-cutting tool distances.

measurements. Moreover, the model can be used to predict the cutting forces over a range of laser scanning and cutting conditions. The force prediction interval is also provided. In this way, the proposed force model can be used to select optimal laser parameters that yield the maximum reduction in cutting forces due to thermal softening.

4.5 References

Aramcharoen, A., Mativenga, P. T., Yang, S., Cooke, K. E., and Teer, D. G. (2008), "Evaluation and selection of hard coatings for micro milling of hardened tool steel," *International Journal of Machine Tools and Manufacture*, 48, 1578-1584.

Bissacco, G., Hansen, H. N., De Chiffre, L. (2005), "Micromilling of hardened tool steel for mould making applications," *Journal of Materials Processing Technology*, 167, 201-207.

Bissacco, G., Hansen, H. N., De Chiffre, L. (2006), "Size effects on surface generation in micro milling of hardened tool steel," *CIRP Annals - Manufacturing Technology*, 55, 593-596.

Budak, E., Altinta, Y., Armarego, E. J. A. (1996), "Prediction of milling force coefficients from orthogonal cutting data," *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 118, 216-224.

Carslaw, H. C., and Jaeger, J. C. (1959), Conduction of heat in solids: Oxford Science Publications.

Guo, Y. B. and Liu, C. R. (2002), "Mechanical properties of hardened AISI 52100 steel in hard machining processes," *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 124, 1-9.

Loewen, E. G. and Shaw, M. C. (1954), "On analysis of cutting-tool temperatures," *Transactions of the ASME*, 76, 217-225.

Marsh, E. R. (2008), Precision spindle metrology: DEStech Publications, Inc..

Melkote, S., Kumar, M., Hashimoto, F., Lahoti, G. (2009), "Laser assisted micro-milling of hard-to-machine materials," *CIRP Annals - Manufacturing Technology*, 58, 45-48.

Mukund Kumar (2011), Laser Assisted Micro Milling of Hard Materials, Ph.D. Thesis of Georgia Institute of Technology.

Singh, R. K and Melkote, S. N. (2005), "Experimental characterization of Laser-Assisted Mechanical Micromachining (LAMM) process," in *2005 ASME International Mechanical Engineering Congress and Exposition, IMECE 2005*, November 5, 2005 - November 11, 2005, Orlando, FL, United states, 2005, 957-964.

Singh, R. K. and Melkote, S. N. (2007), "Characterization of a hybrid laser-assisted mechanical micromachining (LAMM) process for a difficult-to-machine material," *International Journal of Machine Tools and Manufacture*, 47, 1139-1150.

Singh, R. K. and Melkote, S. N. (2007), "Force modeling for laser assisted micro-grooving," *ASME Conference Proceedings*, 2007, 371-379.

Singh, R. K., Joseph, V. R., and Melkote, S. N. (2011), "A statistical approach to the optimization of a laser-assisted micromachining process," *International Journal of Advanced Manufacturing Technology*, 53, 221-230.

Wright, P. K. (1982), "Predicting the shear angle in machining from workmaterial strain-hardening characteristics," *Journal of Engineering for Industry*, 104, 285-292.

CHAPTER V

CONCLUSIONS AND FUTURE RESEARCH

5.1 Conclusions and Original Contributions

This dissertation is the research drawing attention to model enhancement, which widely and significantly exists in various processes and systems. The consideration of the model enhancement has considerable impacts on modeling, design, and optimization of manufacturing processes. This dissertation contributes to model enhancement research by integrating physics-based model and statistical model to mitigate the individual drawbacks and provide models with better accuracy by combining the strengths of both models. The proposed model enhancement methodologies including the following two streams: (1) data-driven enhancement approach and (2) engineering-driven enhancement approach. The methodological developments are generally applicable to various application domains and we have applied the proposed methodologies to (1) nugget diameter prediction of spot welding process, (2) force prediction of laser drilling process, (3) quantification of nanoparticle dispersion of polymer composites, and (4) force prediction and reduction of LAMM process. The original contributions of this dissertation include the following aspects:

- *Model Calibration through Minimal adjustment*: We introduce the minimal adjustment procedure for model enhancement, which brings the physical model closer to the data by making minimal changes to it. This is achieved by a simultaneous variable selection of the model and experimental bias terms. Our procedure starts with a Gaussian Process (GP) model and then approximates it using a linear regression model. The variable selection using nonnegative garrote is then applied to select the important variables in the linear regression model along with the experimental bias terms.

  Two real examples and simulations are presented to demonstrate the advantages of using the minimal adjustment procedure for model enhancement. The minimal adjustment procedure respects the physical model in a better way than

the GP model adjustment, which is an commonly used existing methodology, by acknowledging the possibility of systematic errors in the experiment. Thus, the valuable information in the physical model will not be lost due to a few experimental errors.

- *Quantitative Characterization and Modeling Strategy of Nanoparticle Dispersion in Polymer Composites*: The nanoparticles and clusters are randomly distributed in polymer matrix. Transmission electron microscopy (TEM) images are commonly used to represent the nanoparticle dispersion which plays an essential role in determining the mechanical properties of polymer nanocomposites. Hence, there is a strong need to develop the quantitative measure to effectively describe nanoparticle dispersion state based on microscopy measurement data.

  The number of nanoparticles per unit area at a certain site is proposed to be modeled as a nonhomogeneous Poisson random field. The intensity function of a nonhomogeneous Poisson random field is a mixture of the linear regression model and the Gaussian Markov Random Field (GMRF). To be more specific, the linear regression model is constructed with process variables and nanoparticle characteristics while the GMRF is approximated from Lennard-Jones potential, which is the engineering knowledge of particle interaction force. Unlike the conventional modeling methods that usually rely only on pure physical laws or statistical data-driven techniques; the proposed model integrates both nanomanufacturing domain knowledge and statistical data analysis to provide a better characterization of nanoparticle dispersion. Both simulation and experimental data analysis show the effectiveness of the proposed modeling strategy.

- *Enhancing the Force Model of the Laser Assisted Micro Milling (LAMM) Process* : A physics-based force model for LAMM that captures the effect of thermal softening due to laser heating is required to better understand LAMM process for experimental planning. The developed force model considers runout effect which brings the original force model closer to LAMM in practice. However, there are unknown runout parameters in addition to the unknown model parameter such as friction angle which need to be calibrated simultaneously to achieve

accurate model predictions. We presented a systematic way to estimate these parameters statistically with force measurements collected from the experimentation. Besides, the proposed force model provided the ability to predict the cutting forces over a range of laser scanning and cutting conditions in a timely manner which facilitates the LAMM process optimization to yield the maximal force reduction due to thermal softening. Moreover, the force prediction interval has been derived by incorporating the variability in the runout parameters as well as the variability in the measured cutting forces. The LAMM experiments are carried out on 52100 bearing steel (62 HRc) over a range of feed rates and laser powers. The experimentation result validates that the proposed force model predicts the cutting force profile with good accuracy using a 95% confidence interval.

5.2 Future Research

There are several potential research topics to be explored for further development of model enhancement research. Here are a few examples including both theoretical and practical perspectives.

- *Model Calibration through Minimal adjustment*:
  - New construction of confidence interval: Our approach is frequentist method although it has a Bayesian flavor due to the use of a prior variance-covariance matrix for obtaining the ridge regression estimates, which are used as base estimates for the parameters in the nonnegative garrote. A disadvantage of the frequentist approach is that it does not account for parameter uncertainties as it does in the Bayesian approach. Although this has little effect on the model predictions, the confidence intervals around the predictions can be narrower. One could try to account for all parameter uncertainties to construct new confidence interval for further comparison.
  - Incorporation of prior knowledge: In our approach, we have assumed there is no prior knowledge about the experimental bias. However, in many cases, the experimenter might know when and how they have happened. Consider a situation where the experimenter knows a bias has occurred at certain

setting but does not know about its magnitude (of course, if the magnitude is also known, then it can be easily removed from the data.) This knowledge can be incorporated into the procedure by removing the corresponding θ_i from the linear constraint in the nn-garrote method. Thus, no penalty will be given to the setting where a bias is known to have occurred and therefore it will always be selected in the variable selection procedure. Another scenario could be the experimenter has the confidence/knowledge which there exists experimental biases at a set of possible locations, but don't know the exact ones. In this case, one can try to incorporate such prior knowledge with the proposed minimal adjustment procedure, extend the procedure and make necessary modifications to obtain better predictions.

  - Investigation of large-scale data set: The minimal adjustment procedure is not a mistake-free method. It can sometimes confuse between model and experimental bias terms. However, on an average it is found to perform much better than the GP model adjustment. One can further study that how large the data amount could be incorporated so that the proposed adjustment framework can still distinguish model and experimental bias successfully without failing in the variable selection procedure. Such asymptotic results could be quite interesting and worthy of future study.

- *Quantitative Characterization and Modeling Strategy of Nanoparticle Dispersion in Polymer Composites*:

  - Further investigation of process control and monitoring developments: The proposed nonhomogeneous Poisson random field model has successfully linked process variables with final nanoparticle dispersion state together. It provides a basis for people to further develop process monitoring strategies for nanocomposite production process. Moreover, the root cause diagnosis and active process control developments deserve further investigation as well. The ultimate objective is to achieve uniform nanoparticle dispersion in polymer composites which satisfies the specific mechanical properties required in engineering applications.

- *Enhancing the Force Model of the Laser Assisted Micro Milling (LAMM) Process* :
  - Relaxing Normal assumption on cutting tool axis offset: The actual shift of the tool axis with respect to the spindle axis changes between experiments. This necessitates the need to account for the variability in the predicted cutting forces between different experiments. In this work, we assume that the value of cutting tool axis offset in each run follows Normal distribution and the associated mean and variance parameters could be estimated from experimental data. One could relax this assumption on cutting tool axis offset and develop new approach to incorporate tool axis offset effect into the force model. The parameter estimation schemes would need to be advanced under this new framework.
  - Effects of various types of hard materials: In this work, the LAMM experimentation has been carried out on 52100 bearing steel (62 HRc) and the calibrated force model works well with this material. However, the proposed force prediction model is expected to be generally applicable to other materials. It will be interesting to apply the proposed force model to other kinds of materials, make the appropriate model modifications for such extension, and further compare the robustness of predictive ability over different types of materials.

5.3 References

Chang, C. -J. and Joseph, V. R. (2011), "Model Calibration through Minimal Adjustment," *Technometrics*, tentatively accepted.

Chang, C. -J., Xu, L., Huang, Q., and Shi, J. (2011), "Quantitative Characterization and Modeling Strategy of Nanoparticle Dispersion in Polymer Composites," *IIE Transactions, Special Issue on Quality, Sensing and Pyrognostics Issues in Nanomanufacturing* , in press.

Kumar, M., Chang, C. -J., Joseph, V. R., and Melkote, N. S. (2012), "Laser Assisted Machining Process Modeling and Optimization," to be submitted to *ASME Transactions on Manufacturing Science and Engineering.*

REFERENCES

[1] Aramcharoen, A., Mativenga, P. T., Yang, S., Cooke, K. E., and Teer, D. G. (2008), "Evaluation and selection of hard coatings for micro milling of hardened tool steel," *International Journal of Machine Tools and Manufacture*, 48, 1578-1584.

[2] Balazs, A.C., Emrick, T., and Russell, T.P. (2006), "Nanoparticle Polymer Composites: Where Two Small Worlds Meet," *Science*, 314, 1107-1110.

[3] Bates, D. M., and Watts, D. G. (1988), *Nonlinear Regression Analysis and Its Applications*, New York: Wiley.

[4] Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C. H., Tu, J. (2007), "A Framework for Validation of Computer Models," *Technometrics*, 49, 138–154.

[5] Bissacco, G., Hansen, H. N., De Chiffre, L. (2005), "Micromilling of hardened tool steel for mould making applications," *Journal of Materials Processing Technology*, 167, 201-207.

[6] Bissacco, G., Hansen, H. N., De Chiffre, L. (2006), "Size effects on surface generation in micro milling of hardened tool steel," *CIRP Annals - Manufacturing Technology*, 55, 593-596.

[7] Box, G. E. P. and Hunter, W. G. (1962). "A Useful Method for Model-Building". *Technometrics* 4, pp. 301318.

[8] Breiman, L. (1995). "Better Subset Regression Using the Nonnegative Garrote". *Technometrics*, 37, 373–384.

[9] Budak, E., Altinta, Y., Armarego, E. J. A. (1996), "Prediction of milling force coefficients from orthogonal cutting data," *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 118, 216-224.

[10] Carslaw, H. C., and Jaeger, J. C. (1959), Conduction of heat in solids: Oxford Science Publications.

[11] Chang, C. -J. and Joseph, V. R. (2011), "Model Calibration through Minimal Adjustment," *Technometrics*, tentatively accepted.

[12] Chang, C. -J., Xu, L., Huang, Q., and Shi, J. (2011), "Quantitative Characterization and Modeling Strategy of Nanoparticle Dispersion in Polymer Composites," *IIE Transactions, Special Issue on Quality, Sensing and Pyrognostics Issues in Nanomanufacturing* , in press.

[13] Chapman, R., Mulvaney, P. (2001), "Electro-optical shifts in silver nanoparticle films," *Chemical Physics Letters*, 349, 358 -362.

[14] Cramér, H. and Leadbetter, M. R. (1967), *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*, New York: Wiley.

[15] Deng, X., Joseph, V. R., Mai, W., Wang, Z. L., and Wu, C. F. J. (2009), "A Statistical Approach to Quantifying Elastic Deformation of Nanomaterials," *Proceedings of the National Academy of Sciences*, 106, 11845–11850.

[16] Dobrev, T., Dimov, S. S., and Thomas, A. J. (2006), "Laser Milling: Modelling Crater and Surface Formation," *Proceedings of the Institution of Mechanical Engineers, Part C: J. Mechanical Engineering Science*, 220, 1685–1696.

[17] Efron, B., Johnstone, I., Hastie, T., and Tibshirani, R. (2004),"Least Angle Regression (with discussion)," *The Annals of Statistics*, 32, 407–499.

[18] Fan, J. and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96, 1348-1360.

[19] Guo, Y. B. and Liu, C. R. (2002), "Mechanical properties of hardened AISI 52100 steel in hard machining processes," *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 124, 1-9.

[20] Hamada, M. and Wu, C. F. J. (1992), "Analysis of designed experiments with complex aliasing," *Journal of Quality Technology*, 24, 130-137.

[21] Han, G., Santner, T. J., and Rawlinson, J. J. (2009), "Simultaneous Determination of Tuning and Calibration Parameters for Computer Experiments," *Technometrics*, 51, 464-474.

[22] Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, New York: Springer.

[23] Hench, L.L., West, J.K. (1990), "The Sol-Gel Process," *Chem. Rev.*, 90, 33-72.

[24] Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004), "Combining Field Data and Computer Simulations for Calibration and Prediction," *SIAM Journal of Scientific Computing*, 26, 448–466.

[25] Hildebrand, F. B. (1987), *Introduction to Numerical Analysis*, New York: Dover Publications.

[26] Huang, Q., Wang, L., Dasgupta, T., Zhu, L., Sekhar, P.K., and, Bhansali, S., An, Y., 2010, "Statistical Weight Kinetics Modeling for Silica Nanowires Growth Catalyzed by Pd Thin Film," *IEEE Transactions on Automation Science and Engineering*, DOI: 10.1109/TASE.2010.2070493

[27] Huang, Q. (2011), "Physics-Driven Bayesian Hierarchical Modeling of Nanowire Growth Process At Each Scale," *IIE Transactions*, Vol. 43, pp. 1-11.

[28] Jin, R. and Shi, J., 2011, "Reconfigured Piecewise Linear Regression Tree for Multistage Manufacturing Process Control," *IIE Transactions*, in-press.

[29] Joseph, V. R. (2006). "A Bayesian Approach to the Design and Analysis of Fractionated Experiments," *Technometrics* 48, 219–229.

[30] Joseph, V. R. and Delaney, J. D. (2007). "Functionally Induced Priors for the Analysis of Experiments," *Technometrics* 49, 1–11.

[31] Joseph, V. R. and Melkote, S. N. (2009), "Statistical Adjustments to Engineering Models," *Journal of Quality Technology*, 41, 362–375.

[32] Kennedy, M. C. and O'Hagan, A. (2001), "Bayesian Calibration of Computer Models (with discussion)," *Journal of Royal Statistical Society - Series B*, 63, 425–464.

[33] Kojima Y., Usuki, A., Kawasumi, M., Okada, A., Fukushima, Y., Kurauchi, T., and Kamigaito, O. (1993), "Mechanical properties of nylon 6-clay hybrid," *Journal of Materials Research*, 8, 1185-1189.

[34] Krishnamoorti, R. and Vaia, R.A. (2002), Polymer nanocomposites, 804. Washington, DC: ACS.

[35] Kumar, M., Chang, C. -J., Joseph, V. R., and Melkote, N. S. (2012), "Laser Assisted Machining Process Modeling and Optimization," to be submitted to *ASME Transactions on Manufacturing Science and Engineering*.

[36] Liu, F., Bayarri, M. J., and Berger, J. O. (2009), "Modularization in Bayesian Analysis, with Emphasis on Analysis of Computer Models," *Technometrics* 4, 119–150.

[37] Loewen, E. G. and Shaw, M. C. (1954), "On analysis of cutting-tool temperatures," *Transactions of the ASME*, 76, 217-225.

[38] Mackay, M. E., Tuteja, A., Duxbury, P. M., Hawker, C. J., Horn, B. V., Guan, Z., Chen, G., and Krishnan, R. S. (2006), "General Strategies for Nanoparticle Dispersion," *Science*, 311, 1740-1743.

[39] Marsh, E. R. (2008), Precision spindle metrology: DEStech Publications, Inc..

[40] Maskara, A., Smith, D.M. (1997), "Agglomeration during the drying of fine silica powders .2. The role of particle solubility," *Journal of the American Ceramic Society*, 80(7), 1715-1722.

[41] Matejka, L., Dusek, K., Noga, J. (1998), "Formation, structure and mechanical properties of organic-silica hybrid networks," *Wiley Polym Networks Group Rev Ser*, 1, 301-311.

[42] Melkote, S., Kumar, M., Hashimoto, F., Lahoti, G. (2009), "Laser assisted micro-milling of hard-to-machine materials," *CIRP Annals - Manufacturing Technology*, 58, 45-48.

[43] Mukund Kumar (2011), Laser Assisted Micro Milling of Hard Materials, Ph.D. Thesis of Georgia Institute of Technology.

[44] Pan, E., Ye, L., Shi, J., Chang, T. -S, (2009), "On-line bleeds detection in continuous casting processes using engineering-driven rule-based algorithm", *ASME Transactions, Journal of Manufacturing Science and Engineering*, Volume 131, Issue 6.

[45] Qian, Z. and Wu, C. F. J. (2005), "Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments," *Technometrics*, 50, 192–204.

[46] Qian, Z., Wu, H. and Wu, C. F. J. (2008), "Gaussian Process Models for Computer Experiments with Qualitative and Quantitative Factors," *Technometrics*, 50, 383–396.

[47] Reese, C. S., Wilson, A. G., Hamada, M., Martz, H. F., and Ryan, K. J. (2004), "Integrated Analysis of Computer and Physical Experiments," *Technometrics*, 46, 153–164.

[48] Rubin, D.B. (1980), "Using Empirical Bayes Techniques in the Law School Validity Studies," *Journal of the American Statistical Association*, 75, 801-816.

[49] Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, New York: Springer.

[50] Singh, R. K and Melkote, S. N. (2005), "Experimental characterization of Laser-Assisted Mechanical Micromachining (LAMM) process," in *2005 ASME International Mechanical Engineering Congress and Exposition, IMECE 2005*, November 5, 2005 - November 11, 2005, Orlando, FL, United states, 2005, 957-964.

[51] Singh, R. K. and Melkote, S. N. (2007), "Characterization of a hybrid laser-assisted mechanical micromachining (LAMM) process for a difficult-to-machine material," *International Journal of Machine Tools and Manufacture*, 47, 1139-1150.

[52] Singh, R. K. and Melkote, S. N. (2007), "Force modeling for laser assisted micro-grooving," *ASME Conference Proceedings*, 2007, 371-379.

[53] Singh, R. K., Joseph, V. R., and Melkote, S. N. (2011), "A statistical approach to the optimization of a laser-assisted micromachining process," *International Journal of Advanced Manufacturing Technology*, 53, 221-230.

[54] Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistics Society Series B*, 58, 267–288.

[55] Usuki, A., Kojima, Y., Kawasumi, M., Okada, A., Fukushima, Y., Kurauchi, T., and Kamigaito, O. (1993), "Synthesis of nylon 6-clay hybrid," *Journal of Materials Research*, 8, 1179-1184.

[56] Wang, S., Chen, W., and Tsui, K.-L. (2009), "Bayesian Validation of Computer Models," *Technometrics*, 51, 439–451.

[57] Weng, W.H., Chen, H., Tsai, S.P., Wu, J.C. (2004), "Thermal property of epoxy/SiO2 hybrid material synthesized by the sol-gel process," *Journal of Applied Polymer Science*, 91(1), 532-537.

[58] Wilson, O., Wilson, G.J., Mulvaney, P. (2002), "Laser writing in polarized silver nanorod films," *Advanced Materials*, 14, 1000-1004.

[59] Wright, P. K. (1982), "Predicting the shear angle in machining from workmaterial strain-hardening characteristics," *Journal of Engineering for Industry*, 104, 285-292.

[60] Wu, C. F. J. and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York, NY: Wiley.

[61] Xiong, S. (2010), "Some Notes on the Nonnegative Garrote," *Technometrics*, 52, 349-361.

[62] Xiong, Y., Chen, W., Tsui, K.-L., and Apley, D. (2009), "A better understanding of model updating strategies in validating engineering models," *Computational Methods in Applied Mechanical Engineering*, 198, 1327–1337.

[63] Yoon, PJ., Fornes, T.D., Paul, DR. (2002), "Thermal expansion behavior of nylon 6 nanocomposites," *Polymer*, 43, 6727-6741.

[64] Yuan, M. and Lin, Y. (2007), "On the the nonnegative garrote estimator," *Journal of the Royal Statistics Society-Series B*, 69, 143–161.

[65] Yuan, M., Joseph, V. R., and Zou, H. (2009), "Structured Variable Selection and Estimation," *Annals of Applied Statistics*, 3, 1738–1757.

[66] Zeng, Q.H., Yu, A.B., Lu, G.Q. (2008), "Multiscale modeling and simulation of polymer nanocomposites," *Progress in Polymer Science*, 33, 191-269.

[67] Zhang, H., Zhang, Z., Friedrich, K., Eger, C. (2006), "Property improvements of in situ epoxy nanocomposites with reduced interparticle distance at high nanosilica content," *Acta Materialia*, 54, 1833-1842.

[68] Zhao, H., Jin, R., Wu, S. and Shi, J., 2011, "PDE-constrained Gaussian Process Model on Material Removal Rate of Wire Saw Slicing Process," *ASME Transactions, Journal of Manufacturing Sciences and Engineering*, Vol. 133, 2.

[69] Zou, H. and Hastie, T. (2005), "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society-Series B*, 67, 301-320.